

Treball de Fi de Grau

Grau en Enginyeria en Tecnologies Industrials

Anàlisi dels softwares de visualització de dades i aplicació a l'elaboració de dashboards amb Qlik

MEMÒRIA

Autor: Agustí Oliveres Mallol
Director: Sara Fontdecaba Rigat
Convocatòria: Juny 2017



Escola Tècnica Superior
d'Enginyeria Industrial de Barcelona



Resum

Aquest projecte es basa en la creació d'una eina de suport per a la presa de decisions empresarials centrades en evitar la pèrdua de clients per tal de maximitzar el seu temps de vida dins l'empresa.

Inicialment es fa un estudi comparatiu entre els diferents softwares d'intel·ligència de negocis. Es determinen les diferents funcions a valorar en els softwares i s'avalua cada una d'aquestes funcions pels softwares més importants del mercat actual.

A continuació, s'aprofundeix en el problema empresarial de la pèrdua de clients, estudiant quines estratègies apliquen les empreses, quins paràmetres s'utilitzen per a parametritzar i controlar els clients o factors que inciten la marxa del client, entre altres.

Finalment, es du a terme la creació d'una eina informàtica per a la gestió de la pèrdua de clients utilitzant un dels esmentats programes d'intel·ligència de negocis. Es simula tot el procés que es podria trobar una empresa i es detallen tots els passos des de la captura d'informació fins a la creació final de l'esmentada eina.

Sumari

RESUM	2
SUMARI	3
ÍNDEX DE FIGURES	6
ÍNDEX DE TAULES	8
1. GLOSSARI	9
1.1. Nomenclatura	9
2. PREFACI	10
2.1. Origen del projecte	10
2.2. Motivació	10
2.3. Requeriments previs	10
3. INTRODUCCIÓ	12
3.1. Objectius del projecte	12
3.2. Abast del projecte	12
4. CONTEXTUALITZACIÓ	13
5. SITUACIÓ ACTUAL DEL MERCAT DE BI SOFTWARES	14
6. COMPARATIVA ENTRE SOFTWARES	15
6.1. Qlik Sense	16
6.2. Tableau	18
6.3. Microsoft Power BI	20
6.4. Resum	22
7. EXPLICACIÓ DEL CAS PRÀCTIC DEL PROJECTE	23
8. INTRODUCCIÓ I DESCRIPCIÓ DEL <i>CHURN</i> I DEL <i>CUSTOMER LIFETIME VALUE</i>	24
8.1. <i>Churn</i> de clients com a problema empresarial	24
8.2. Cicle de vida d'un client (<i>Customer lifetime</i>)	26
8.3. <i>Customer Lifetime Value</i> (CLV)	28
9. ANÀLISI DEL <i>CHURN</i>	29
9.1. Punt de partida	29
9.2. Base de dades	29
9.3. Taules i els seus camps	31
9.3.1. Botigues	31

9.3.2.	SegParameters	32
9.3.3.	Dictionary	33
9.3.4.	Dicpattern	34
9.3.5.	Shopweekday	34
9.3.6.	History	35
9.3.7.	Customer	35
9.3.8.	Clientchar	37
9.3.9.	AeaOut	38
9.3.10.	ClvOut	39
10.	EINES QUE S'UTILITZARAN	41
10.1.	MySQL	41
10.2.	Qlik Sense	42
11.	IMPLEMENTACIÓ A QLIK SENSE	43
11.1.	Disseny dels dashboards	43
11.2.	Tractament i transformació de les dades (ETL)	44
11.2.1.	Taules finals	45
11.2.1.1.	Taula Base	46
11.2.1.2.	Taula Alarm_CP_NP	49
11.2.1.3.	Taula History	51
11.2.2.	Relació entre les taules finals	51
11.2.3.	Taules addicionals	52
11.3.	Creació de la interfície interactiva	55
11.3.1.	Customer Granularity	57
11.3.2.	Wrap Up	58
11.3.3.	General	59
11.3.4.	General_2	60
11.3.5.	At Risk	61
11.3.6.	At Risk_2	62
11.3.7.	Customer Overview	64
11.3.8.	Customer Detail	65
11.3.9.	Customer Detail_2	66
11.3.10.	Map	67
11.4.	Ús de l'aplicació	67
12.	EXECUCIÓ DEL PROJECTE	69
12.1.	Tasques principals	69
12.2.	Planificació detallada	70
13.	COST ASSOCIAT	71
13.1.	Despeses directes	71
13.2.	Despeses indirectes	72

13.3. Despeses totals	72
14. IMPACTE AMBIENTAL	74
CONCLUSIONS	75
AGRAÏMENTS	77
BIBLIOGRAFIA	78
Referències bibliogràfiques	78
ANNEXES	79
A1. Càrrega de les taules a Qlik Sense	79
A1.1. Secció Main	79
A1.2. Secció Map	80
A1.3. Secció Wrap Up.....	80
A1.4. Secció General	81
A1.5. Secció Date Filters.....	83
A1.6. Secció Variables	83
A1.7. Secció At Risk NP.....	84
A1.8. Secció History	85

Índex de Figures

<i>Figura 1 “Màgic quadrant” dels diferents softwares de BI on es distribueixen segons la seva posició en el mercat segons la consultora Gartner.</i>	14
<i>Figura 2 Diferents maneres de crear valor i les diferents accions que pot fer una empresa per augmentar el nombre de clients ([8]).</i>	25
<i>Figura 3 Diferents etapes del cicle de vida d'un client. Es representa el valor generat pel client en funció del temps, així com els aspectes a procurar de millorar per cada etapa. A més a més, es distingeix entre 3 tipus de clients diferents: gold, silver, bronze ([8]).</i>	26
<i>Figura 4 Conjunt d'etapes que segons la consultora Bain & Company passa el client abans de cometre el Churn, evidenciant que no és degut a un fet puntual, sinó a una acumulació de circumstàncies</i>	27
<i>Figura 5 Flux d'informació entre les diferents taules de la base de dades proporcionada per Datancia.</i>	30
<i>Figura 6 Flux de treball que s'ha dut a terme pel procés de ETL.</i>	45
<i>Figura 7 Representació esquemàtica dels diferents tipus de joins. Cada cercle representa un taula i la zona colorejada el resultat final de la join en qüestió.</i>	46
<i>Figura 8 Representació de la combinació de dues taules a partir de dos columnes de referència</i>	48
<i>Figura 9 Representació de l'ús d'una taula (Botigues) com a diccionari per afegir noves dimensions a una taula original.</i>	48
<i>Figura 10 Esquema de les taules utilitzades per a la creació de la taula Base amb les corresponents dimensions. Les que es troben en el mig són les dimensions comunes a cada taula i que s'han utilitzat com a paràmetres d'enllaç.</i>	49
<i>Figura 11 Esquema del flux de treball utilitzat per a l'obtenció de la dimensió evolution_NP</i>	50
<i>Figura 12 Esquema del model de dades carregat a Qlik Sense. Es veu com les tres taules finals es relacionen entre elles per la clau sintètica customerID+period.</i>	52
<i>Figura 13 Esquema de les taules addicionals carregades a Qlik Sense, on es veu que s'enllacen entre elles per la dimensió customerID_WUP</i>	53
<i>Figura 14 Esquema final del conjunt de taules carregades a Qlik Sense i com s'enllacen entre elles</i>	54
<i>Figura 15 Visió de la App creada a l'interfície de Qlik Sense</i>	55
<i>Figura 16 Visió del conjunt de dashboards dins de la App a Qlik Sense</i>	56

<i>Figura 17 Visió del dashboard Customer Granularity</i>	57
<i>Figura 18 Visió del dashboard Wrap Up</i>	58
<i>Figura 19 Visió del dashboard General</i>	59
<i>Figura 20 Visió del dashboard General_2</i>	60
<i>Figura 21 Visió del dashboard At Risk</i>	61
<i>Figura 22 Representació del càlcul dels diferents paràmetres dels gràfics del dashboard</i>	62
<i>Figura 23 Visió del dashboard At Risk_2</i>	62
<i>Figura 24 Visió del dashboard Customer Overview</i>	64
<i>Figura 25 Visió del dashboard Customer Detail</i>	65
<i>Figura 26 Visió del dashboard Customer Detail_2</i>	66
<i>Figura 27 Visió del dashboard Map</i>	67
<i>Figura 28 Diagrama de Gantt que mostra el temps requerit per cada tasca del projecte</i>	70

Índex de Taules

<i>Taula 1 Comparativa dels diferents criteris per cada software</i>	22
<i>Taula 2 Mostra de la taula Botigues</i>	32
<i>Taula 3 Mostra de la taula SegParameters</i>	32
<i>Taula 4 Mostra de la taula Dictionary</i>	33
<i>Taula 5 Mostra de la taula Dicpettern</i>	34
<i>Taula 6 Mostra de la taula Shopweekday</i>	34
<i>Taula 7 Mostra de la taula History</i>	35
<i>Taula 8 Mostra de la taula Customer</i>	36
<i>Taula 9 Mostra de la taula Clientchar</i>	37
<i>Taula 10 Mostra de la taula AeaOut</i>	38
<i>Taula 11 Mostra de la taula ClvOut</i>	40
<i>Taula 12 Disseny de què es mostrarà, com es mostrarà, i per a qui aniran destinats els dashboards</i>	43
<i>Taula 13 Dimensions de la taula final Base</i>	47
<i>Taula 14 Significat dels diferents valors de la dimensió evolution_NP</i>	50
<i>Taula 15 Dimensions que fomen la taula final Alarm_CP_NP</i>	51
<i>Taula 16 Dimensions que formen la taula final History</i>	51
<i>Taula 17 Dimensions de les taules addicionals Wrap_Up_History i Wrap_Up_Customer</i>	53
<i>Taula 18 Dimensions que formen la taula addicional Date</i>	54
<i>Taula 19 Mostra de les diferents dimensions de la taula Date per un period donat.</i>	54
<i>Taula 20 Conjunt de dimensions disponibles en els filtres</i>	57
<i>Taula 21 Significat de cada un dels KPIs que mostra el dashboard</i>	58
<i>Taula 22 Tasques principals del projecte</i>	69
<i>Taula 23 Cost humà del projecte en funció del preu per hora i les hores totals invertides</i>	71
<i>Taula 24 Costos associats a recursos materials</i>	71
<i>Taula 25 Cost total de les despeses directes</i>	72
<i>Taula 26 Despeses indirectes</i>	72
<i>Taula 27 Detall de les despeses totals del projecte</i>	72

1. Glossari

1.1. Nomenclatura

BI: *Business Intelligence*

IT: *Information Technology*

ETL: *Extraction Transformation and Loading*

KPI: *Key Performance Indicator*

API: *Application programming interface*

CLV: *Customer Lifetime Value*

PaaS: *Platform as a Service*

SaaS: *Software as a Service*

BBDD (DB): Bases de dades (*DataBase*)

AEA: *Activitat Estimada Actual*

GUI: *Graphical User Interface*

ODBC: *Open Database Connectivity*

SQL: *Structured Query Language*

2. Prefaci

El projecte neix de la necessitat actual en què es troben moltes empreses de disposar d'una eina informàtica per tractar de forma amena, visual i sobretot interactiva, tota la informació de què disposen. En altres paraules: crear valor útil per a l'empresa a partir de dades. Concretament, s'ha centrat en el problema de la fidelització i pèrdua de clients.

En aquest treball s'abarca: els principals programes informàtics que existeixen actualment per transformar els volums d'informació de les empreses en formats de més fàcil interpretació, el problema empresarial de la pèrdua de clients i els paràmetres principals per mesurar-lo, i per últim, el cas pràctic de l'anàlisi de la pèrdua de clients d'una empresa utilitzant un dels programes informàtics anteriors.

2.1. Origen del projecte

Aquest projecte neix de la mà de l'empresa Datancia, que ha jugat un paper protagonista en la realització d'aquest treball. Datancia es trobava en la situació de prendre una decisió sobre quin programa informàtic d'intel·ligència de negocis utilitzar, i amb la intenció de donar suport i facilitar aquesta decisió, va sorgir la necessitat i els objectius del present treball, realitzant un estudi comparatiu dels diferents softwares, i duent a terme un cas pràctic a partir d'una base de dades proporcionada per Datancia mateix.

2.2. Motivació

La motivació principal d'aquest projecte ha estat poder oferir una solució a un problema real en què es troben moltes empreses actualment. Tant és així, que les pràctiques curriculars com a estudiant d'aquesta universitat, les he fet en el departament de *Business Analysis* en l'empresa líder de classificats de l'estat.

Així doncs, el tractament de la informació de forma òptima per tal d'extreure'n valor ha estat un tema que sempre m'ha interessat i aquest treball m'ha permès involucrar-me en tot el procés que suposa el tractament i l'anàlisi de les dades, amb tot l'aprenentatge sobre processos, terminologia i llenguatges informàtics que implica.

2.3. Requeriments previs

Per a la realització de la part pràctica d'aquest projecte s'ha hagut de tractar amb un

conjunt de programes informàtics específics, amb diferents llenguatges de programació. Per sort, en el meu cas ja havia tractat amb aquests llenguatges anteriorment, i aquest treball m'ha permès aprofundir-hi i dominar-los encara més.

Deixant de banda els coneixements sobre llenguatges informàtics, ha estat de gran ajuda comptar amb coneixements bàsics sobre bases de dades relacionals, com s'estructuren, i com interactuar amb elles.

Per últim, també ha estat imprescindible comptar amb una base de dades d'una empresa, la qual en aquest cas, ha estat proporcionada gràcies a l'empresa Datancia.

3. Introducció

3.1. Objectius del projecte

Els objectius principals que pretén respondre aquest projecte són els següents:

- Donar una visió detallada de l'estat actual del mercat sobre programes informàtics referents a la intel·ligència de negocis.
- Identificar els programes líders del mercat, explicar els punts forts i febles de cada un, i comparar-los entre ells.
- Aprofundir sobre el problema empresarial de la pèrdua de clients: com les empreses encaren aquest problema, l'evolució d'un client, paràmetres per avaluar el risc de perdre un client...
- Crear una plataforma visual i interactiva de dades que faciliti la presa de decisions empresarials.
- Detallar, pas a pas, tot el procés de creació d'una plataforma interactiva per analitzar i gestionar la pèrdua de clients d'una empresa, partint de les dades recopilades per la mateixa.

3.2. Abast del projecte

En primer lloc es fa una introducció teòrica i detallada de tots els aspectes relacionats amb els programes sobre la intel·ligència de negoci i el problema de la pèrdua de clients, on s'aprofundeix en els programes més importants i es detallen els conceptes clau del problema empresarial.

En segon lloc, fent referència a la part pràctica del projecte, es du a terme tota la creació de la interfície gràfica de les dades des del més començament, partint de les dades brutes en forma de taula, tractant-les i transformant-les de manera adequada, i finalment l'elaboració de la interfície amb un dels programes detallats prèviament.

4. Contextualització

No seria erroni afirmar que vivim en la societat de la informació. Centrant-nos en el món empresarial, gràcies a Internet i al desenvolupament dels sistemes d'informació, les empreses i els seus directius disposen cada cop de molta més informació, de més qualitat i a una velocitat trepidant.

Aquest augment de disponibilitat de dades ofereix un gran potencial per tal de millorar la presa de decisions, dur a terme els objectius proposats i detectar noves oportunitats de negoci. No obstant això, les empreses topen amb el següent problema: tenir cada cop més informació i menys temps i eines per analitzar-la. Aquest problema es pot convertir en un autèntic repte de gestió per a les empreses, les quals busquen disposar de capacitat per prendre decisions amb rapidesa, basades en coneixement real del mercat i de les seves tendències, per tal de disposar d'una font d'avantatge competitiu.

A arrel d'aquest problema, neix el concepte de *Business Intelligence* (BI), definit de la següent manera pel llibre *Research and Practical Issues of Enterprise Information Systems* [1]:

“BI consisteix en el conjunt d'estratègies, processos, aplicacions, informació, tecnologies i architectures tècniques usades per recollir, analitzar, presentar i divulgar informació empresarial.”

En altres paraules, el principal objectiu de la BI és donar suport de forma continuada a les organitzacions per facilitar la presa de decisions i millorar la seva competitivitat. Per tal de complir aquests objectius, es disposen dels *BI Softwares*: programes informàtics destinats a organitzar i analitzar la informació. Entre altres funcionalitats d'aquests softwares, com ara l'extracció i “neteja” de dades provinents de diferents fonts d'informació per tal de poder-les explorar i analitzar posteriorment, la principal funció pel que es valoren aquests programes és per l'elaboració d'eines visuals i interactives d'exploració de la informació (*visual-based data discovery*), també anomenades Dashboards.

En el mercat actual s'hi poden trobar una gran varietat de *BI softwares*. Alguns més enfocats a l'extracció de les dades de les diferents fonts que es puguin tenir, altres al tractament i “neteja” d'aquestes dades i altres a eines de *reporting*, que inclou eines visuals com els comentats *dashboards*. En aquest treball ens centrarem en l'estudi i comparació d'alguns dels diferents productes de *Business Intelligence*, enfocats a l'elaboració de *dashboards*.

5. Situació actual del mercat de BI softwares

El mercat de plataformes de BI *visual-based data discovery* va començar als voltants de l'any 2004 amb productes destinats a experts i tècnics en tecnologies de la informació (IT) de les diferents organitzacions. Actualment es busca que aquests productes siguin accessibles a un públic més extens, amb eines més senzilles i fàcils perquè un major nombre d'usuaris el puguin utilitzar i treure'n profit.

Hi ha un gran nombre de BI softwares en el mercat actual. Segons la consultora tecnològica *Gartner* [2], el mercat actual presenta varis factors que evidencien la maduresa del propi mercat, com l'augment de la demanda per part dels compradors de noves i millors prestacions dels productes, la pressió sobre els preus i la disminució de diferències entre els competidors. Tot i així, entre tots aquests competidors, tant l'empresa *Gartner* com la majoria dels entesos en el sector, estan d'acord en què hi ha tres competidors que lideren el mercat (Figura 1): Microsoft Power BI [3], Qlik [4] i Tableau [5].

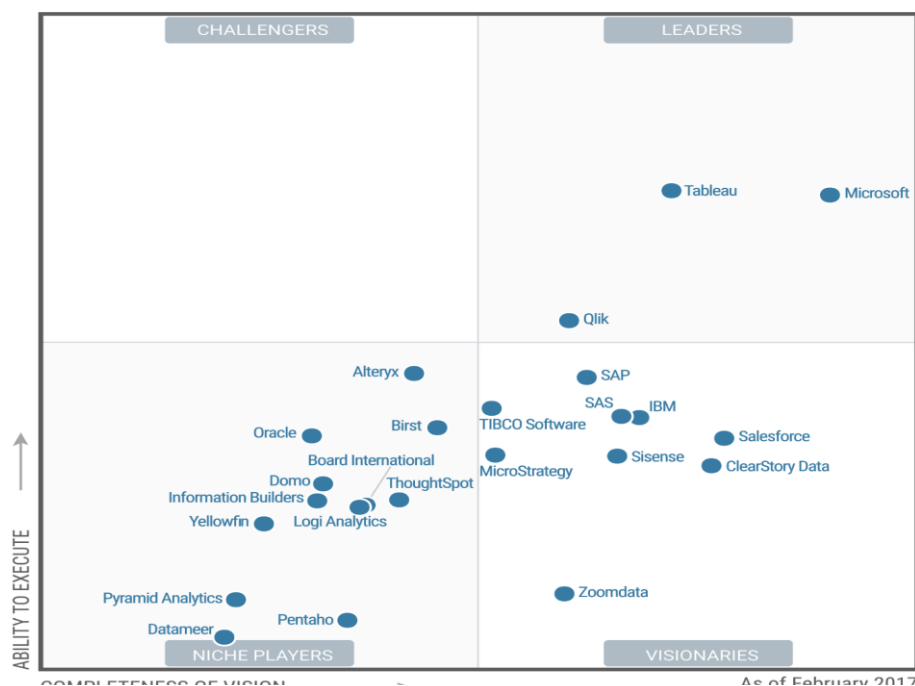


Figura 1 "Màgic quadrant" dels diferents softwares de BI on es distribueixen segons la seva posició en el mercat segons la consultora *Gartner*.

La següent part del projecte tracta sobre els tres softwares BI líders del mercat. Es realitza un anàlisi intensiu de les fortaleeses i punts febles de cada un d'ells, i finalment es conclou amb una comparativa directa entre ells.

6. Comparativa entre softwares

Per tal de comparar els diferents softwares, seguint els criteris establerts principalment per *Gartner* (en el seu informe *Magic Quadrant for Business Intelligence and Analytics Platforms* [6], Març 2017) amb petites modificacions segons criteris propis i de diferents entesos involucrats, s'han definit una sèrie de premisses que es puntuaran per cada plataforma.

- Infraestructura
 - BI en el núvol (*cloud BI*): poder treballar sense necessitat d'un programa d'escriptori amb totes les eines implementades correctament¹.
 - Administració i seguretat de la plataforma: poder administrar els accessos dels usuaris, recuperació de dades, compatibilitat amb diferents sistemes operatius, etc.
 - Connectivitat i “ingesta” de dades de diferents fonts: permetre la connexió a informació estructurada i desestructurada², des de diferents fonts d'emmagatzematge i des de fonts *online* o *on-premises*³.
- Facilitat per transformar les dades (*data management*)
 - ETL⁴ i emmagatzematge de dades propis: poder accedir, transformar i integrar les dades.
 - Gestió de la informació. Oferir de forma centralitzada i robusta el poder buscar, emmagatzemar i utilitzar les dades per obtenir mesures, indicadors (KPIs), eines visuals, etc.
 - Preparació intel·ligent: oferir capacitats avançades d'unions entre diferents taules (*intelligent joins*, *hierarchy generation*...).
- Anàlisis i creació de contingut

¹ Conceptes relacionats: PaaS i SaaS.

² Informació estructurada presenta un alt grau de d'organització i categorització, facilitant operacions algorítmiques i el seu tractament. En canvi, aquella desestructurada li manca aquesta organització.

³ Fonts *on-premises* fan referències a aquelles tecnologies localitzades físicament a la pròpia empresa/organització. Per contra, les fonts *online* s'executen remotament des de servidors externs o des del “núvol”.

⁴ Sigles de: Extraction, Transformation and Loading.

- Eines d'anàlisis integrades. Permetre a l'usuari l'accés a eines d'anàlisi d'alt nivell integrades a la pròpia plataforma, o importades de desenvolupadors externs.
- Exploració visual i interactiva: oferir un gran nombre d'opcions de visualització, així com poder manipular les dades de forma interactiva amb els gràfics.
- “Descobriments intel·ligents”. Es refereix a fer descobriments de forma automàtica, com correlacions o prediccions sense la necessitat d'aplicar algorismes o models manualment.
- Implementació per a dispositius mòbils. Permetre visualitzar i desenvolupar continguts als dispositius mòbils, així com aprofitar les seves capacitats, com la pantalla tàctil, la càmera o la localització.
- Compartir contingut
 - Integració a altres plataformes. Oferir la capacitat d'incloure el software a altres suports o aplicacions, així com APIs⁵ detallades.
 - Compartir i publicar els descobriments. Oferir capacitats com permetre als usuaris publicar el seu contingut, compartir pensaments o debatre sobre els resultats, o també l'opció de crear una alarma quan succeeixi un canvi determinat en els valors.
- Aspectes generals
 - Bona integració de les necessitats (*workflow*): consisteix en avaluar fins a quin grau totes les capacitats que ofereix el programa estan implementades en un sol producte o en múltiples productes amb dificultats per relacionar-se entre ells.
 - Facilitat d'ús i atracció visual: avaluar la senzillesa a l'hora de crear contingut visual, administrar accessos, etc.
- Preu

6.1. Qlik Sense

L'empresa de software Qlik presenta dos softwares diferents: Qlik Sense i Qlik View. Sense és el nou producte que s'ofereix als nous clients i l'aposta principal de l'empresa per al

⁵ Application Programming Interface (API). Conjunt d'indicacions perquè una funció o procediment d'un programa en concret pugui ser utilitzat i interaccionar amb un altre programa

futur, mentre que Qlik View és l'antic software que un gran nombre dels seus clients encara manté i utilitza. Encara hi ha certes prestacions de View que no estan implementades a Sense, per això en alguns casos encara s'han de mantenir els dos programes, però la intenció per part de l'empresa és aconseguir una transició completa de View cap a Sense.

Fortaleses:

- Aplicacions robustes. Qlik Sense permet ser utilitzat com un pseudo *data mart*⁶. Permet acceptar varies fonts d'informació i càlculs complexos amb aquestes dades. Presenta grans resultats a l'hora de treballar amb un gran volum de dades en memòria (*in-memory*).
- Facilitat d'ús i atracció visual. Es destaca la facilitat d'implementar dimensions i relacionar-les entre elles de forma senzilla a través de la seva plataforma. El seu joc de colors permet veure de forma ràpida els filtres usats, així com els valors exclosos i actius en cada filtre.
- Implementació de dashboards de Qlik en altres aplicacions. Presenta una API molt detallada perquè altres clients puguin implementar els seus dashboards a les seves aplicacions.

Àrees a millorar:

- Manca de processos analítics avançats. Processos predictius o de *clustering*⁷ no estan disponibles encara en el software. Tampoc està implementada la funció per poder cridar i executar scripts de R⁸ o Python⁹, tot i que aquesta funció ja ha estat confirmada per la companyia que per a la següent actualització del software estarà disponible.
- Limitacions en aplicacions mòbil. Tot i que es pot utilitzar el software Qlik Sense en dispositius mòbils, no existeixen "apps" exclusives per aquests dispositius, pel que no s'aprofiten funcionalitats com la localització o notificacions *push*. En el seu pla de futur, també està previst que es solucioni aquesta manca.

⁶ Magatzem de dades (data warehouse) a petita escala centrat en un tema o una àrea de negoci específic d'una organització (departamental)

⁷ Tècnica molt comuna en l'anàlisi estadística que es basa en la classificació i agrupació d'objectes similars en diferents subconjunts (o clústers).

⁸ Entorn i llenguatge de programació de software lliure, vastament utilitzat, enfocat a l'anàlisi estadístic.

⁹ Llenguatge de programació molt popular, que es basa en la seva senzillesa sintàctica.

- Funcionalitats en el servei *cloud* i altres mancances. Presenta mancances pel que fa a funcionalitats administratives i de seguretat requerides per les empreses. No presenta la funció d'enllaçar dades provinents d'Internet i pròpies del sistema físic (*on-premises*). No presenta una forma fàcil per a usuaris inexperts de crear expressions entre les dimensions del dashboard.

Preus:

- QlikSense
 - Desktop: Gratuït.
 - Enterprise: \$1500/TOKEN. 1 TOKEN = Ús il·limitat per 1 usuari o 10 *Logins* de temps limitat. El segon està pensat per usuaris que accedeixen puntualment o poc freqüentment.
- QlikSense Cloud
 - Cloud Basic: Gratuït. Permet usar totes les funcions del producte, i compartir-ho fins a 5 usuaris.
 - Cloud Plus: \$20/mes. Permet compartir amb un nombre il·limitat d'usuaris, així com una capacitat d'emmagatzematge superior.
 - Cloud Business: \$25/usuari/mes. Permet treballar a nivell de grups, editar espais de treball simultàniament, i automatitzar càrregues d'aplicacions de forma automàtica.

6.2. Tableau

Tableau presenta un producte altament interactiu i visualment intuïtiu. Ofereix tres productes diferents: Tableau Desktop, Tableau Server i Tableau Online¹⁰ (servei *cloud*), i està preparat per poder treballar i desplaçar-se pels tres productes diferents de forma fàcil i senzilla. S'ha centrat principalment en usuaris de caràcter "no tècnic", el que li ha repercutit amb un gran creixement des de la seva arribada al mercat. Actualment, però, degut a la pròpia maduresa del mercat, s'ha vist forçat a implementar característiques orientades a tècnics en IT.

Fortaleses:

¹⁰ Desktop és l'aplicació d'escriptori bàsica de disseny. Tableau Online representa l'opció basada en el "núvol", i el producte Server representa l'opció de les prestacions on-premises. Aquests dos últims productes són usats per poder compartir contingut.

- Exploració i interacció visual intuïtiva. La fortalesa principal de Tableau són les funcions d'exploració i anàlisis intuïtives en els dashboards. Permet visualitzar dades de pràcticament qualsevol font, sent la creació de contingut visual molt senzilla, així també com la formulació d'expressions complexes entre les dimensions a representar.
- Integrar funcions provinents de R o *Python*, podent incorporar algorismes dins de càlculs de Tableau.
- Accés a dades provinents d'una gran varietat de fonts diferents. Permet treballar amb dades provinents de la pròpia memòria del sistema (*in-memory*), estructurades o desestructurades (Hadoop, NoSQL), d'arxius individuals personals, de paquets estadístics (SAS, R, etc.), o de fonts *on-line*.
- Suport a dispositius mòbils. El contingut de Tableau està implementat i totalment interactiu amb dispositius *Android* o *iOS*.

Àrees a millorar:

- Treballar amb varies *fact tables*¹¹. Tableau no contempla treballar en un mateix projecte amb dades de diferents *fact tables*, encara que provinguin de la mateixa font. També presenta limitacions a l'hora de reutilitzar i treballar amb camps o segments calculats prèviament.
- Preparació de les dades. La plataforma no suporta transformacions o retocs de les dades. El programa fa de forma automàtica algunes accions molt bàsiques com “despivotar” taules o eliminar cel·les buides provinents de fulls de càlcul, però no permet canvis més complexes. Això provoca que les dades ja han d'estar preparades i “cuinades” originalment en el moment de l'extracció de la informació de les diferents fonts. Està previst que surti una nova eina que permeti aquesta transformació i preparació de les dades.
- Mancances sobre implementacions i exportacions en altres plataformes. Tableau no està implementada en Linux, i presenta moltes limitacions a l'hora d'implementar les seves eines en altres programes (APIs poc elaborades). Tampoc estan disponibles, encara, funcions com exportacions a PDF o PowerPoint i alertes.
- Baix rendiment amb dades provinents de la pròpia memòria del sistema. Tableau no presenta tan bons resultats com els seus competidors en extraccions d'un gran nombre de dades en memòria.

¹¹ Taula central d'un esquema de dades dimensional. Aquesta taula té després associades varies taules, cada una d'una dimensió (o categoria) diferent.

Preus:

- Tableau Desktop
 - Personal: \$35/usuari/mes. Permet connectar-se a dades d'arxius guardats en el propi sistema (fulls de càlcul, Access, fitxers estadístics, fitxers de text...) i guardar de forma local.
 - Professional: \$70/usuari/mes. Permet connectar-se a pràcticament qualsevol tipus de font, i guardar projectes localment o publicar-los *on-line* a Tableau Server o Tableau Online.
- Tableau Server: \$35/usuari/mes. Permet publicar en espais de treball interns o públics amb Amazon Web Services, Microsoft Azure i Google Cloud Platform. És l'espai on poder manejar la informació i els dashboards realitzats, i poder-hi accedir directament des del navegador o el dispositiu mòbil.
- Tableau Online: \$42/usuari/mes. És el producte de Tableau Server, però entregat de forma *SaaS*¹², amb manteniment, actualitzacions i millores periòdiques.

6.3. Microsoft Power BI

Eina que forma part del servei de Microsoft: *Microsoft Azure*¹³. L'empresa, que també és la proveïdora del full de càlcul *Excel*, ofereix serveis de BI integrats (*add-ins*) a aquest com: Power Query, Power Pivot, Power View i Power Map. Power BI compta amb la plataforma on-line, tot i que si es vol tractar informació provinent de fonts *on-premises*, l'aplicació "Desktop" (instal·lada en el propi sistema, en memòria) és necessària. Microsoft acostuma a millorar el software amb modificacions mensuals.

Fortaleses:

- Facilitat d'ús i atracció visual. El producte destaca per ser un dels més senzills d'utilitzar. A més a més de la pròpia interfície, facilita la feina a l'usuari ja que Microsoft ofereix la infraestructura del "magatzem" i la transformació de dades, així com per la publicació dels dashboards.

¹² Software as a Service (SaaS). És un model de distribució de softwares on el suport lògic i la informació del programa s'allotgen en servidors d'una companyia externa, on el client hi accedeix a través d'internet. Per tant, no requereix una instal·lació del software en el sistema informàtic del client.

¹³ Col·lecció de múltiples serveis IT en el núvol que ofereix Microsoft

- *Quik Insights*. Opció del software que automàticament crea els gràfics més significatius per les dades que se li dona.
- Preparació de les dades i bon rendiment *in-memory*. Permet transformar i netejar les dades a mesura que es van carregant a l'aplicació. Permet un alt rendiment al tractar dades provinents de la pròpia memòria del sistema, podent tractar un gran nombre de dades (tot i que a l'aplicació del *cloud* l'espai està limitat a 10 GB per usuari).
- Accessibilitat a un gran nombre de fonts d'informació diferents i aplicacions predefinides. Microsoft presenta un excel·lent suport a la majoria de tipus de fonts de les dades (tant estructurades com desestructurades). A més, compta amb un conjunt d'aplicacions predefinides que faciliten el tractament de dades, la creació de models i de gràfics.

Aspectes a millorar:

- Funcions analítiques avançades. Power BI no ofereix opcions avançades d'anàlisis integrades en el propi menú de la plataforma, tot i que sí ofereix l'opció de cridar un *script* de R des de la plataforma Desktop. També hi ha l'opció d'integrar visualitzacions més avançades des del propi *Marketplace* de Microsoft.
- No compta amb eines importants bàsiques de visualització. Sorprenentment, Power BI encara no ofereix l'opció de crear taules dinàmiques (*pivot tables*), ni de crear taules amb subtotals acumulats. Pot ser degut a que es vulgui que es creïn primer amb *Excel* i després s'importin al dashboard, però continua sent un procés poc pràctic.
- L'opció per augmentar la capacitat d'emmagatzematge no està clara ni ben definida. Per elaborar anàlisis complexos es necessita d'altres plataformes, com *Excel*, *Azure Machine Learning* o R, creant un *workflow*¹⁴ poc ben definit i ineficient.

Preus:

- Power BI Free: Gratuït. Està limitat a 1GB d'emmagatzematge per usuari i a 10 mil files/hora en quant a capacitat de consum de dades. La freqüència per poder actualitzar contingut de manera automàtica és diària. Els usuaris poden compartir el contingut.
- Power BI Pro: \$9,99/usuari/mes. Capacitat de 10GB d'emmagatzematge i capacitat de 1 milió de files/hora. Ofereix més prestacions en quant a accedir a diferents fonts d'informació.

¹⁴ Flux de treball

6.4. Resum

La següent taula (Taula 1) avalua els diferents criteris establerts anteriorment. Les avaluacions estan basades en l'informe dut a terme per la consultora *Gartner*, on els resultats estan recolzats per una combinació entre tot un conjunt d'enquestes i qüestionaris sobre els diferents productes tant a venedors com a compradors i usuaris dels softwares, i les opinions dels seus propis analistes.

Les avaluacions són en una escala de l'1 al 5, on:

1	Absent o molt pobre
2	Pobre
3	Correcte
4	Excel·lent
5	Sobrepassa els requeriments

Taula 1 Comparativa dels diferents criteris per cada software

Categoria	Criteris	Qlik	Tableau	Microsoft Power BI
Infraestructura	Cloud BI	3	3	3,5
	Administració i seguretat de la plataforma	4	3,5	4
	Connectivitat i ingesta de dades	3	4	4,5
Data Management	ETL i emmagatzematge	5	3	4
	Gestió de la informació	3	3	3
	Preparació intel·ligent	3	3	3
Anàlisi i creació del contingut	Eines d'anàlisis integrades	2	3	2
	Exploració visual i interactiva	3,5	4	4
	Descobriments intel·ligents	2	2	2,5
	Dispositius mòbils	2,5	4	4
Compartir contingut	Compartir i publicar descobriments	3	2	2
	Integració a altres plataformes	4,5	3	3,5
Aspectes generals	Workflow	3	5	2,5
	Facilitat d'ús i atracció visual	4	4,5	4
Preu		3	2	4

7. Explicació del cas pràctic del projecte

El problema que es vol arribar a solucionar és la manca d'una eina d'anàlisi i gestió de la pèrdua de clients d'un supermercat. Es parteix del cas on una empresa es troba amb la necessitat de controlar la pèrdua dels seus clients, identificant aquells que es troben en risc d'abandonar l'empresa i el cost econòmic associat a aquesta pèrdua, amb l'objectiu de minimitzar les baixes de forma efectiva.

El cas pràctic consistirà, llavors, en tot el viatge des de les dades brutes que recull la pròpia empresa sobre l'activitat dels seus clients i la transformació d'aquestes en paràmetres que serveixin per analitzar l'evolució de l'activitat dels consumidors, fins a la creació d'una plataforma interactiva que permeti la visualització i el tractament de totes les dades i paràmetres per tal d'obtenir resultats que aportin valor real i ajudin a la presa de decisions de l'empresa.

Les diferents etapes i processos que es duran a terme en la part pràctica d'aquest projecte, ordenades per ordre cronològic, són les següents:

1. Obtenció de les dades sobre l'activitat dels clients provinents de l'empresa de supermercats. En aquest cas, les dades són simulades i obtingudes de part de l'empresa Datancia¹⁵ [7].
2. Transformació de les dades a través de l'algorisme dissenyat per Datancia, per tal de parametritzar l'activitat dels clients i obtenir indicadors sobre el risc de perdre un client.
3. Tractament de totes les dades obtingudes per tal d'obtenir una base de dades centralitzada i amb tota la informació relacionada entre ella.
4. Disseny i creació del *dashboard* sobre tota la informació obtinguda, per tal de donar tot el suport necessari per gestionar la pèrdua de clients i identificar les opcions de negoci adequades.

¹⁵ Empresa que obté i genera informació i valor a partir de les dades dels seus clients, implementant solucions i tecnologies de vanguardia que generen una major eficiència en les seves pràctiques de gestió.

8. Introducció i descripció del *Churn* i del *Customer Lifetime Value*

Els actuals processos de globalització i el cada vegada més gran volum de dades que es disposen per analitzar estan canviant els marcs competitius en la majoria dels sectors econòmics, deixant enrere aquelles companyies que no són capaces d'innovar i aprofitar les noves oportunitats de negoci que apareixen cada cop de manera més freqüent. L'aparició de nous competidors i noves tecnologies implica un fort augment de la competència, i una especial preocupació per a les empreses del conegut sector terciari o de serveis.

Per tal d'obtenir avantatge competitiu davant la competència, cada cop més empreses decideixen canviar (o reduir) el seu objectiu de captar com més nous clients millor, i centrar-se en mantenir i retenir els clients que ja tenen, amb la intenció d'anticipar-se i evitar el fenomen conegut com *churn*, és a dir, la pèrdua general de clients.

Degut a la necessitat d'identificar aquests clients, neixen tot un conjunt de processos estadístics i nous conceptes, com el *customer lifetime value*, per tal d'identificar patrons i categoritzar l'activitat dels clients.

8.1. *Churn* de clients com a problema empresarial

En l'escenari actual, anteriorment descrit, d'augment continu de la pressió i competència entre empreses, el procés de creació de valor per a l'empresa i el seu desenvolupament passa per retenir i prolongar la vida dels seus clients, així com augmentar el seu consum mitjà (Figura 2).

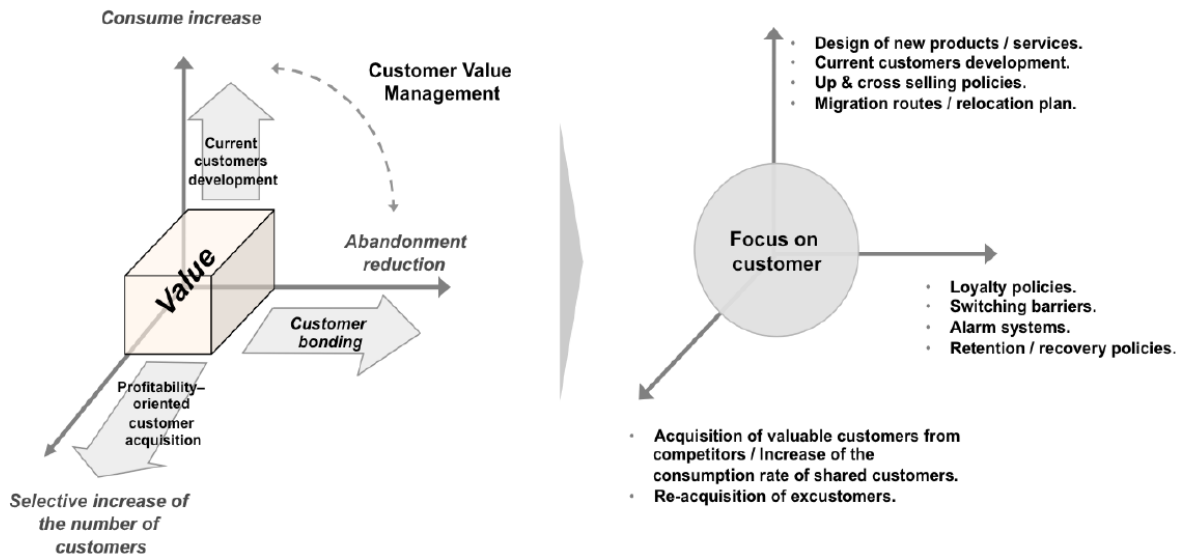


Figura 2 Diferents maneres de crear valor i les diferents accions que pot fer una empresa per augmentar el nombre de clients ([8]).

D'aquesta manera, una estratègia comercial "defensiva", orientada a mantenir i crear "enllaços de lleialtat" amb el client existent, és molt més eficient i menys costosa que una estratègia agressiva basada en l'expansió de la seva posició en el mercat i intentar captar el màxim nombre de clients. Es converteix així en una prioritat entendre els mecanismes d'augment de la fidelitat del client, junt amb el llançament d'accions per retenir-lo.

Com a conseqüència, és normal que el paradigma comercial de grans empreses estigui canviant de captura massiva de nous clients cap a la conservació dels existents. Tot i això, les empreses no ho tenen fàcil per crear fidelitat entre els clients, ja que aquests estan exposats contínuament a un gran nombre d'ofertes anunciades i promocionades per la competència. Actualment el coneixement i l'exigència dels clients dels diferents mercats creix constantment, fet que provoca que la creació d'enllaços de fidelitat es converteixi en un factor clau per a la continuació de la companyia en el mercat.

Les companyies de serveis tenen com a principal i més important actiu el client. No només per ser els usuaris i consumidors dels seus productes/serveis, sinó perquè un client satisfet es converteix en ambaixador de la teva marca al públic. Per això, tenen la responsabilitat d'implementar polítiques comercials per permetre desenvolupar el màxim potencial comercial del client. És a dir, prolongar el màxim la vida activa del client en l'empresa, prenent les mesures adequades per cada etapa del seu "cicle de vida" (Figura 3).

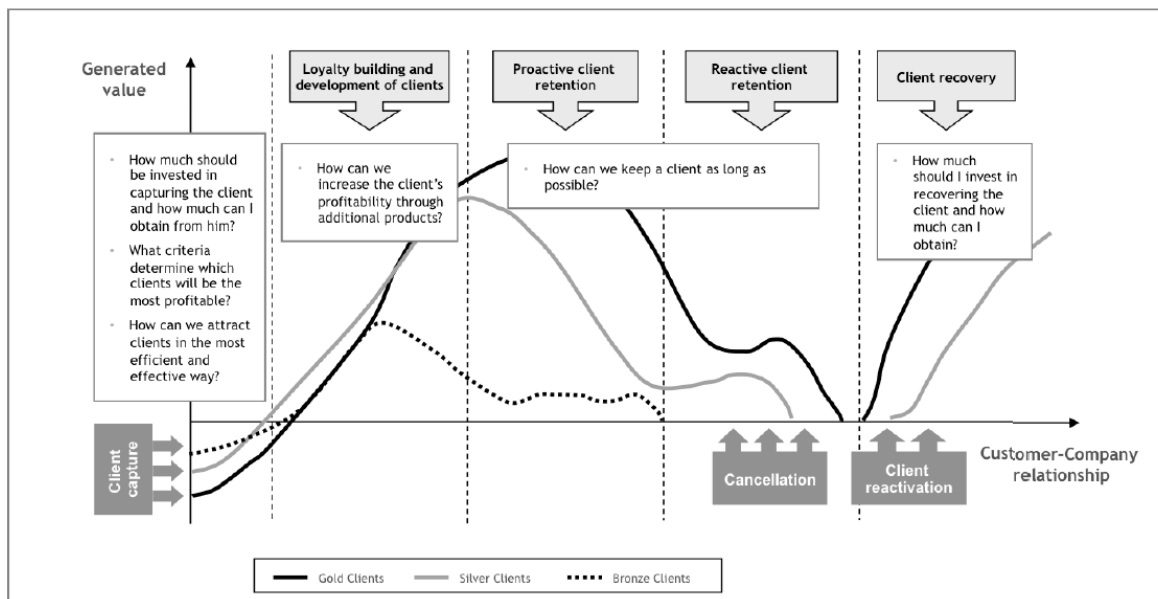


Figura 3 Diferents etapes del cicle de vida d'un client. Es representa el valor generat pel client en funció del temps, així com els aspectes a procurar de millorar per cada etapa. A més a més, es distingeix entre 3 tipus de clients diferents: *gold*, *silver*, *bronze* ([8]).

Augmentar l'esperança de vida del client hauria de ser la base de qualsevol acció comercial amb l'objectiu d'aportar i crear valor per a l'empresa. Adoptant estratègies a curt termini, orientades a incrementar els nivells de venda dels clients en comptes de respondre i afavorir a les seves necessitats, pot conduir a mals resultats en horitzons de temps més llargs, provocant descontentament i insatisfacció que inevitablement afavoriria al *churn*.

8.2. Cicle de vida d'un client (*Customer lifetime*)

L'anteriorment comentat *cicle de vida* d'un client és un concepte vital per als departaments de *marketing* per tal de maximitzar l'eficiència de les accions, tàctiques i estratègies comercials. Es sol definir aquest cicle amb cinc etapes diferents:

- 1) Adquisició. Fase on el cost de les inversions necessàries és major que els recollits provinents dels clients. S'inverteix en la creació de bases de dades (BBDD) per poder analitzar els comportaments i obtenir noves oportunitats.
- 2) Conversió. Etapa on el client ha tingut alguna mena de contacte amb l'empresa o el producte, i les accions estan orientades a captar el client.
- 3) Creixement. S'ha aconseguit captar el client. El principal objectiu és el de adequar les accions per tal d'aconseguir el màxim potencial de cada client, i mantenir-lo tant de temps com sigui possible.

- 4) Retenció. L'objectiu principal a aquest punt és procurar que el client continuï consumint i que no se'n vagi a la competència.
- 5) Reactivació. Una vegada el rendiment del client a cessat o està apunt de fer-ho. Els possibles camins per al consumidor és desaparèixer de l'empresa, o bé tornar a aportar valor, on dependrà en gran part de les mesures preses per tornar a captar el client.

És llavors, de vital importància mantenir el màxim temps possible l'estat actiu i el màxim potencial del client, així com detectar amb la suficient antelació aquells clients que estan reduint la seva activitat per així concentrar les accions i estratègies pertinents per evitar el *churn*.

Cal també comentar el fet que un gran nombre de directius i executius apunten a un o pocs fets a la pèrdua de clients, com pot ser el llançament d'una campanya publicitària de la competència, l'obertura d'un nou local al costat del seu, etc. Tal com apunta *Bain & Company* [9] en el seu reportatge sobre el *churn*, anomenat *Breaking the back of customer churn* [10], la majoria de les pèrdues de clients en empreses del sector de serveis són degudes a un cúmul de males experiències en l'empresa que finalment acaba fent que el client sigui fàcilment convençut d'abandonar la companyia (Figura 4). Per això, grans empreses compten i inverteixen en polítiques per contentar al client i així afavorir a la seva lleialtat a la companyia.

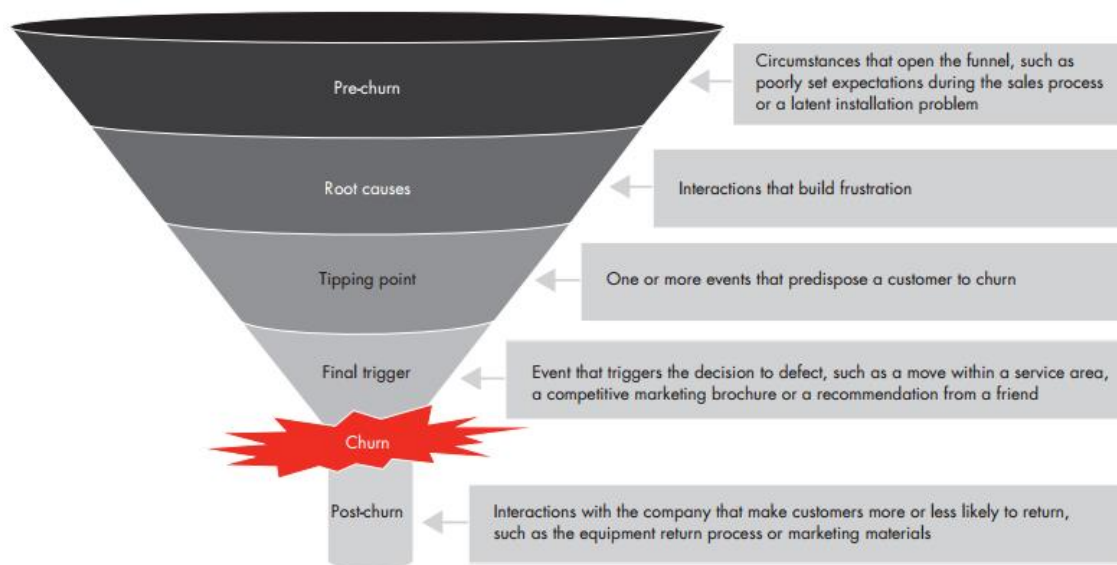


Figura 4 Conjunt d'etapes que segons la consultora *Bain & Company* passa el client abans de cometre el Churn, evidenciant que no és degut a un fet puntual, sinó a una acumulació de circumstàncies

8.3. *Customer Lifetime Value (CLV)*

La definició de la fidelitat d'un client, així com la seva modelització, ha estat tema de discussió durant varis anys en el món del *customer management*¹⁶. La proposta que sembla més acceptada i estesa és la que es basa en la detecció d'aquells clients en risc, categoritzant-los en varies categories en funció de la variació de les seves transaccions comercials. Això implica que la fidelitat s'entén des d'un punt de vista centrat en el client i no en el producte. Per això, s'utilitzarà el CLV, definit com el valor de futurs ingressos provinents d'un client.

Així, aquells clients que disminueixen el seu CLV seran els que estaran en risc de churn, i el seu càlcul correcte permetrà detectar-los a temps i elaborar les estratègies i tàctiques adequades per corregir l'activitat del client.

El CLV també és un factor clau a tenir en compte a l'hora de fer decisions en diferents àmbits empresarials:

- Marketing: Quant puc invertir per captar un nou client?
- Producte: Com puc oferir productes/serveis adaptats als millors clients?
- Suport al client: Quant puc invertir en retenir aquells clients en risc?
- Ventes: Quin tipus de client ens hauríem de centrar en adquirir?

¹⁶ Terme que fa referència a la gestió del client

9. Anàlisi del Churn

9.1. Punt de partida

Per a aquest projecte es disposa d'una base de dades proporcionades per Datancia, que simulen dades sobre ventes i clients amb què podrien comptar una empresa dedicada a la venda al públic, en aquest cas: un supermercat.

Aquestes dades “brutes”, sense cap transformació prèvia, no aporten una visió directa sobre quins clients estan en perill d'abandonar l'empresa o quins ingressos es deixarien de guanyar si es perdés un client en concret. Per tal d'aconseguir aquesta informació, s'aplica un algorisme dissenyat per Datancia per aconseguir tots aquests paràmetres necessaris per analitzar el churn. L'algorisme en qüestió, a través de models estadístics i d'un històric de registres dels clients, s'arriba a obtenir paràmetres com la predicció del que comprarà un client en el futur, la probabilitat que el client es consideri “viu”, o el grau de probabilitat que el client abandoni l'empresa, entre altres.

L'algorisme creat per Datancia és només aplicable a empreses que comptin amb un sistema de seguiment dels seus clients, com poden ser targetes client per seguir la pista de la seva activitat, i que a més a més faci un mínim de dos anys que estigui establert, ja que el model estadístic s'alimenta d'un registre històric de dades que com més ampli sigui, més precisos i acurats seran els resultats, i el mínim temps que assegura la veracitat de les prediccions és de dos anys. És important comentar que aquest algorisme està destinat a empreses d'ús freqüent, com supermercats, grans magatzems o gimnasos, on el consum sigui periòdic.

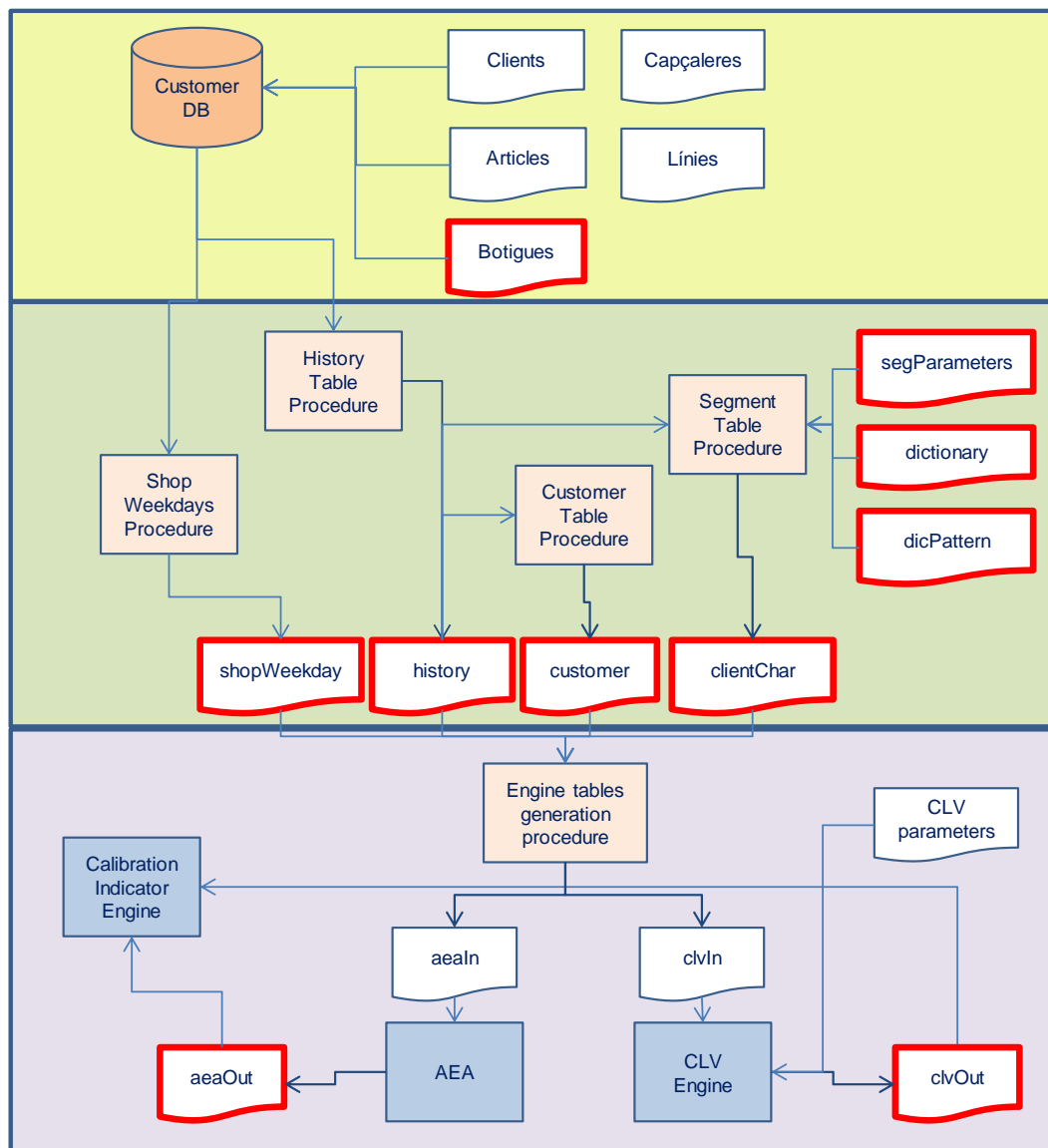
A partir de les noves dades provinents de l'algorisme de modelització del churn, s'analitzaran els resultats per tal d'extreure'n valor útil per afrontar i encarar futures estratègies per impedir la pèrdua de clients. Per aconseguir-ho, es treballarà amb softwares de BI enfocats a l'elaboració de *dashboards*.

9.2. Base de dades

Les dades de les que es disposen són tant les dades originals recollides per l'empresa com les dades finals tractades i passades per l'algorisme d'anàlisi del churn. En aquest cas, com ja s'ha comentat, les dades que provindrien de l'empresa són simulades.

En el següent esquema (Figura 5) es mostra el flux d'informació i el paper que juga cada

taula de la base de dades en el procés:



DATANCIA
MAKING SENSE OUT OF YOUR DATA

Figura 5 Flux d'informació entre les diferents taules de la base de dades proporcionada per Datancia.

En l'esquema anterior s'ha dividit el flux de la informació i les taules que intervenen en el procés en tres etapes, diferenciades en els tres colors de fons diferents. En ordre descendent:

1. Inicialment es parteix de les dades provinents de l'empresa en format de varies taules, creant totes elles la base de dades anomenada "Customer DB" (en aquest



cas, dades simulades). És a partir d'aquestes taules que es creen totes les altres. D'aquestes taules només es disposa de la taula *Botigues*. Aquesta taula és l'única que prové d'una font d'origen diferent a tota la resta, fet que comportarà que s'hagi de tractar diferent a la resta, com s'explica més endavant (11.2.1.1).

2. A continuació, es mostra el procés de tractament de les dades d'entrada provinents del client. Consisteix en estandarditzar i segmentar la informació de tal manera que estigui en el format adequat per aplicar-hi l'algorisme d'anàlisi del churn. Les tres taules d'entrada d'aquesta segona etapa: *segParameters*, *Dicpattern* i *Dictionary*, són les que contenen la informació necessària per segmentar els clients segons la seva activitat històrica. Aquestes taules poden estar basades en mètodes estadístics sobre el mercat dels productes que es tracten, en nocions i paràmetres establerts per la mateixa empresa client, o un híbrid de les dues anteriors.
3. Per últim, s'aplica l'algorisme d'anàlisi del churn desenvolupat per Datancia a les taules finals de l'etapa anterior. S'aconsegueix així, per cada client i per cada període, el nivell d'alarma sobre la probabilitat que el client abandoni la companyia o la predicció del CLV, entre d'altres (9.3).

Les taules d'informació s'han representat de color blanc, i les taules de les que es disposen són aquelles envoltades de color vermell. Per últim, els blocs rectangulars indiquen procediments o motors de creació i modificació de noves estructures de dades, com per exemple, els algorismes de predicció de paràmetres sobre el churn.

9.3. Taules i els seus camps

Com s'ha mostrat en l'esquema del flux d'informació (Figura 5), les taules de les que es disposen es poden separar entre les que intervenen en el procés de tractament i segmentació de les dades inicials, i entre les que s'obtenen després d'aplicar els algorismes sobre el churn. A més a més, es disposa de només una taula provinent de l'empresa en qüestió: la taula *Botigues*.

9.3.1. Botigues

Aquesta taula (Taula 2) aporta informació geogràfica de cada una de les botigues de l'empresa:

Taula 2 Mostra de la taula Botigues

Ciutat	shopId	Longitud	Latitud
Barcelona	38	2,155782	41,392544
Madrid	3	3,694431	40,435122
...

- Ciutat. Nom de la ciutat on es troba la botiga.
- shopId. Codi de referència de la botiga.
- Longitud/Latitud. Coordenades geogràfiques d'on es troba exactament la botiga.

9.3.2. SegParameters

La taula SegParameters (Taula 3) alimenta al procediment (*procedure*¹⁷) encarregat de segmentar els clients segons la seva activitat, en funció dels diferents indicadors. Aquest procediment, anomenat *Segment Table Procedure*, intervé en la creació de la taula ClientChar. La taula SegParameters ha de ser recalculada i revisada fora del cercle d'operacions un cop l'any.

Taula 3 Mostra de la taula SegParameters

IndicatorId	Thr1	Thr2	Thr3	Thr4	Thr5
1	29472.01	6645.27	1757.67	464.36	122.14
2	334.00	210.40	105.69	52.85	26.18

- indicatorId. Cada Id fa referència a un indicador diferent de l'activitat.
 1. overallRevenue: import (€) gastat durant la setmana.
 2. noOfVisits: nombre de visites que el client ha comprat.
- Thr1. Valor per sota del qual l'activitat de l'indicatorId es considera de nivell A.
- Thr2. Valor per sota del qual l'activitat de l'indicatorId es considera de nivell B.
- Thr3. Valor per sota del qual l'activitat de l'indicatorId es considera de nivell C.

¹⁷ Procediment informàtic que permet aplicar lògiques a les taules per tal d'obtenir noves dades a partir de les inicials.



- Thr4. Valor per sota del qual l'activitat de l'indicatorId es considera de nivell D.
- Thr5. Valor per sota del qual l'activitat de l'indicatorId es considera de nivell E.

9.3.3. Dictionary

Per cada combinació de segments categoritzats a la taula SegParameters, assigna un valor de patró del client. Aquesta taula (Taula 4) també s'usa per alimentar el procediment *Segment Table Procedure*, i ha de ser recalculada i revisada fora del cercle d'operacions un cop l'any.

Taula 4 Mostra de la taula Dictionary

segAvgTcktVal	segNoOfVis	segRet	segDiscounts	segFam1	...	segFam10	activityPattern	tracking
A	A	E	E	E	...	E	Ultra High	1
A	B	E	E	E	...	E	Ultra High	1
A	C	E	E	E	...	E	High	1
A	D	E	E	E	...	E	High	1
A	E	E	E	E	...	E	Medium	1
B	D	E	E	E	...	E	Medium	1
B	E	E	E	E	...	E	Low	1
D	C	E	E	E	...	E	Low	1
D	D	E	E	E	...	E	Ultra Low	1
D	E	E	E	E	...	E	Ultra Low	1
...

- segAvgTcktVal. Segment per al preu del ticket.
- segNoOfVis. Segment per al nombre de visites.
- segRet. Segment per al nombre de productes retornats per setmana.
- segDiscounts. Segment per als descomptes aplicats.
- segFamX. Segment per a productes categoritzats com a família X.
- activityPattern. Segment global característic per a l'activitat del client en funció dels segments anteriors.
- Tracking. 0 si el client no és monitoritzat i 1 si és monitoritzat.

9.3.4. Dicpattern

Relaciona cada patró d'activitat categoritzat en la taula Dictionary (camp activityPattern) amb varis indicadors. S'utilitza també per alimentar el procediment *Segment Table Procedure* (Taula 5).

Taula 5 Mostra de la taula Dicpettern

indicattorPattern	activityPattern	Revenue	NoOfVisits
1	Ultra High	Molt alt	Molt alt
2	High	alt	alt
3	Medium	Mig	Mig
4	Low	Baix	Baix
5	Ultra Low	Molt baix	Molt baix

- indicatorPattern. Indicador del patró d'activitat del client.
- activityPattern. Patró d'activitat del client. Coincideix amb el mateix camp de la taula Dictionary.
- Revenue. Patró d'activitat sobre l'indicador de l'import gastat pel client.
- NoOfVisits. Patró d'activitat sobre l'indicador del nombre de visites del client.

9.3.5. Shopweekday

A partir de les dades proveïdes per la companyia i a través del procediment *Shop Week Procedure*, es crea aquesta taula (Taula 6) amb els dies hàbils per cada període i botiga.

Taula 6 Mostra de la taula Shopweekday

shopId	week	openDays
59	1623	6
59	1624	5
59	1625	6
43	1623	6
43	1624	7

- shopId. Codi de referència de la botiga.
- week. Setmana en format YYWW, on els dos primers nombres indiquen l'any i els dos últims la setmana. En totes les taules, la granularitat temporal és a nivell setmanal.



- openDays. Dies hàbils per a la botiga i setmana corresponents.

9.3.6. History

El procediment *History Table Procedure* crea la taula History (Taula 7), a partir també de les dades provinents de l'empresa client. Cada fila representa l'activitat d'un indicador (IndicatorId) per període i client.

Taula 7 Mostra de la taula History

customerId	period	shopId	indicatorId	indicatorValue
50018	1521	40	1	62.08
50018	1527	40	1	61.96
700033	1540	36	1	54.82
890004	1605	44	1	53.78
970031	1607	28	1	0.48
1000024	1605	17	2	6.00
700033	1525	42	2	6.00
890004	1520	13	2	5.00
...

- customerId. Codi únic d'identificació del client.
- period. Període codificat en format any+setmana.
- shopId. Codi de referència de la botiga
- indicatorId. Cada Id fa referència a un indicador diferent de l'activitat.
 1. overallRevenue: import (€) gastat durant la setmana.
 2. noOfVisits: nombre de visites que el client ha comprat.
- indicatorValue. Valor corresponent a cada indicatorId (Preu, nombre de visites, etc.).

9.3.7. Customer

És la taula mestra de clients, que inclou la informació rellevant de cada un d'ells actualitzada periòdicament, segons quin sigui el període mínim d'estudi (dia, setmana, mes...). La taula (Taula 8) és fruit del procediment *Customer Table Procedure*, que a la vegada, aquest s'alimenta del ja comentat *History Table Procedure*, com es veu a l'esquema (Figura 5).

Taula 8 Mostra de la taula Customer

customerId	Birth Date	Street	no	Postal Code	Sex	T0	reg Shop	Tinit	Recency	→
290001	1949	Córsega	3	08008	M	2014-04-27	35	2015-05-18	2016-03-07	→
790002	1983	Balmes	4	08006	F	2014-11-13	45	2015-06-01	2016-03-21	→
410003	1957	Bogatell	5	08005	M	2014-06-14	35	2015-05-18	2016-03-21	→
890004	1990	Diagonal	23	08029	F	2014-12-21	6	2015-05-11	2016-03-21	→
950005	1994	Bogatell	5	08005	F	2015-01-13	42	2015-05-11	2016-03-21	→
...	→

→	trackingType	mainShop	secondaryShop	dayOfWeek	timeOfDay
→	3	35	11	3	4
→	3	45	36	7	1
→	3	35	14	7	2
→	3	6	12	6	1
→	3	42	30	3	3
→

- customerId. Codi únic d'identificació del client.
- birthDate. Any de naixement del client.
- Street. Carrer de residència.
- no. Número de casa de l'adreça del client.
- Postal Code. Codi postal.
- Sex. Sexe del client.
- T0. Període en el que es va donar d'alta el client.
- regShop. Codi de la botiga on el client es va donar d'alta.
- Tinit. Període a partir del qual es comença a calcular el període T (paràmetre que s'usa per a la modelització de l'activitat del client i definit a la taula Clvout) en els casos que T0 sigui massa antic.
- Recency. Data de l'última transacció.



- trackingType. 0: sense activitat rellevant en el període; 1: sense temps d'activitat suficient per a la segmentació (52 setmanes); 2: inclòs en la segmentació; 3: inclòs en la monitorització.
- mainShop. Codi de la botiga habitual on el client realitza la majoria de les seves compres.
- secondaryShop. Codi de la botiga no habitual on el client també realitza compres.
- dayOfWeek. Dia de la setmana que el client acostuma a fer les compres.
- timeOfDay. Franja horària que el client acostuma a fer les compres.
 - 0: matí, entre les 9 i 12 hores.
 - 1: migdia, entre 12 i 15 hores.
 - 2: tarda, entre 15 i 18 hores.
 - 3 vespres, entre 18 i 21 hores.

9.3.8. Clientchar

La taula (Taula 9) prové del procediment *Segment Table Procedure*, que aquest s'alimenta del *History Table Procedure* i les tres taules comentades anteriorment (segParameters, Dictionary i DicPattern). ClientChar consisteix en una taula on per cada client i període es mostren els valors acumulats dels indicadors, així com els diferents segments a què pertany el client.

Taula 9 Mostra de la taula Clientchar

customerId	period	Revenue	noOfVisits	returns	discounts	Revenue Fam1	Seg Revenue	**Dic**
110395	1618	407.40	60	0.00	0.00	0.00	C	...
500310	1618	406.75	79	0.00	0.00	0.00	C	...
220413	1618	406.18	59	0.00	0.00	0.00	C	...
790500	1618	405.67	64	0.00	0.00	0.00	C	...
490807	1618	405.07	76	0.00	0.00	0.00	C	...
...

- customerId. Codi únic d'identificació del client.
- period. Període codificat en format any+setmana.

- Revenue. Import gastat acumulat. Cada nou període (setmana) s'eliminarà el valor associat a la primera setmana i es canviarà pel nou període entrant.
- noOfVisits. Número de vegades que el client a comprat.
- Discounts. Import dels descomptes acumulats pel client.
- revenueFamX. Import gastat acumulat en els productes de la família X.
- ****Dic****. Mateixos camps que la taula Dictionary amb les mateixes definicions.

9.3.9. AeaOut

Juntament amb la taula ClvOut són les taules resultants de l'algorisme d'anàlisi del churn dissenyat per Datancia. La taula AeaOut (Activitat Estimada Actual) és el *DataSet*¹⁸ de sortida del motor de càlcul AEA, que requereix de només una entrada: "aealn". La taula AeaOut (Taula 10) mostra per cada client monitoritzat, període i indicador, tota una sèrie de paràmetres estadístics per determinar i categoritzar si l'activitat del client és l'esperada, millor o pitjor.

Taula 10 Mostra de la taula AeaOut

customerId	period	indicatorId	indicatorValue	baseline	aea	sdAea	lambda	score	alarm
290001	1616	1	0.00	3.88	1.02	0.54	0.25	-1.62	-1
790002	1616	1	0.00	1.33	0.97	0.29	0.10	-0.58	0
410003	1618	1	9.56	1.02	1.69	0.35	0.10	0.82	0
890004	1625	1	9.81	7.42	11.28	0.34	0.15	1.13	1

- customerId. Codi únic d'identificació del client.
- period. Període codificat en format any+setmana.
- indicatorId. Cada Id fa referència a un indicador diferent de l'activitat.
- indicatorValue. Valor corresponent a cada indicatorId (Preu, nombre de visites, etc.)
- baseline. Valor esperat pel client de l'indicatorId basat en la seva activitat històrica.
- aea. Paràmetre indicador de l'activitat actual del client.

¹⁸ Col·lecció d'informació habitualment tabulada



- sdAea. Desviació estàndard del paràmetre aea.
- lambda. Paràmetre de suavització del model estadístic.
- score. Representa la diferència entre la baseline i la aea en base la sdAea: $(\ln(aea + 1) - \ln(baseline + 1))/sdAea$. Indica la variació entre el valor real i l'esperat: si és 0, el real i l'esperat són iguals; si és major que zero, el real és superior a l'esperat; i si és negatiu, el real és inferior a l'esperat.
- alarm. Nombre enter basat en l'score, on indica de forma categòrica si l'activitat del client ha millorat, empitjorat o s'ha mantingut. Seran, llavors, aquells clients amb alarma negativa que en el període en qüestió es podran categoritzar com a clients en risc de churn.

9.3.10. ClvOut

La taula ClvOut (Taula 11), juntament amb Aeaout, és una de les taules resultants de l'algorisme d'anàlisi del churn, i prové del motor de càlcul amb nom *CLV Engine*. Aquest motor de càlcul requereix d'uns paràmetres del model que es calculen la primera vegada, i es reajusten anualment, i que s'emmagatzemen a la taula "CLV parameters". El *CLV Engine*, requereix de dos *dataSet* d'entrada: "clvIn" i "CLV parameters"; per tal d'obtenir-ne un de sortida: "clvOut". Tant "clvIn" com "clvOut" tenen les mateixes columnes, amb la diferència que les de "clvIn" els valors corresponen al període anterior que els de "clvOut", que corresponen al període que s'està analitzant.

El motor de càlcul *CLV Engine* té tres objectius:

1. Determinar la probabilitat que un client realitzi una compra en el futur, i per tant, la probabilitat que el client hagi deixat de comprar.
2. Determinar el volum de compra previst per un client en un període futur predefinit.
3. Determinar el valor residual¹⁹ d'un client per a l'empresa.

La taula mostra una fila per client (monitoritzat o no) i per període, i amb les següents dimensions:

¹⁹ Valor actual net del flux de compres futures del client en qüestió

Taula 11 Mostra de la taula ClvOut

customerId	period	x	t	T	m	Palive	CLVpred	Xpred	Zpred	→
290001	1606	47	35	38	4.79	0.98	428.60	63	6.75	→
790002	1606	19	34	36	5.06	1.00	249.88	29	8.71	→
410003	1606	21	38	38	4.85	1.00	249.18	30	8.35	→
...	→

→	x5	x25	x75	x95	z5	z25	z75	z95
→	1	149	58	238	0.99	4.47	2.95	6.75
→	1	45	29	68	1	5.32	3.05	8.71
→	1	45.3	30	69	1	5.16	3.70	8.35
→

- customerId. Codi únic d'identificació del client.
- period. Període codificat en format any+setmana.
- x. Número de compres realitzades fins al període.
- t. Número de períodes entre la primera compra del client i l'última.
- T. Número de períodes entre la primera compra del client i el període actual.
- m. Import mig de les compres realitzades fins al moment.
- Palive. Probabilitat de que el client estigui "viu".
- CLVpred. *Customer Lifetime Value* (8.3) fins les properes 52 setmanes.
- Xpred. Número de compres que haurà acumulat el client durant les properes 52 setmanes.
- Zpred. Import que haurà acumulat el client durant les properes 52 setmanes.
- xN. Número de compres que haurà acumulat el client durant les properes 52 setmanes amb una funció de probabilitat de N%.
- zN. Import que haurà acumulat el client durant les properes 52 setmanes amb una funció de probabilitat de N%.



10. Eines que s'utilitzaran

Per tal de tractar i analitzar les dades s'ha necessitat de softwares específics especialitzats en dur a terme tasques concretes del procés. Les dues tasques que han necessitat de programari exclusiu per a fer-les han estat:

- Visualització i tractament de les diferents taules de les que es disposen.
- Creació de *dashboards* interactius per graficar de forma senzilla i útil tota la informació que es desitgi.

Per a la primera tasca, s'ha emprat el software lliure *MySQL* ([11]) que fa la funció de servidor de base de dades. Per a la segona tasca s'ha utilitzat un dels softwares de BI, comentats anteriorment (apartat 6), líders del mercat. D'entre els tres productes que lideren el mercat actual de *Business Intelligence* (Qlik, Tableau i Power BI) s'ha escollit treballar amb *Qlik*. Els motius de la decisió s'expliquen en el proper apartat 10.2.

10.1. MySQL



S'ha emprat el software lliure de *Oracle Corporation*²⁰: *MySQL* versió 5.7.18.1, que, com ja s'ha comentat, fa la funció de servidor de base de dades i permet gestionar, modificar, unir i un sens fi d'altres opcions sobre les taules de partida. Tot aquest tractament es fa amb el llenguatge de programació SQL, fet que ha comportat un procés d'aprenentatge individual del llenguatge per a realitzar totes les consultes i transformacions necessàries.

Per a utilitzar *MySQL* d'una manera més amena i assequible per a nous usuaris, també s'ha fet ús de l'extensió *MySQL Workbench Visual Database Designer 6.3*, que ofereixen els mateixos proveïdors del programa principal i també de forma gratuïta. El *MySQL Workbench* és el la interfície gràfica d'usuari (GUI²¹) oficial per a *MySQL*, que permet

²⁰ Empresa nord-americana, líder en el seu sector, que desenvolupa programari per gestionar bases de dades electròniques i productes relacionats. Alguns dels seus productes més populars són: *MySQL*, *Oracle Database* i *Java*.

²¹ *Graphical User Interface*. Interfície d'usuari que utilitza elements gràfics, sonors i de control per facilitar la interacció amb un sistema informàtic de manera més intuïtiva que no pas el clàssic sistema per línia d'ordres, més difícil d'aprendre i dominar.

dissenyar, crear, consultar, i navegar en els esquemes de bases de dades a través d'una interfície més senzilla i que facilita la gestió de la informació.

Per últim, i per tal de poder connectar *MySQL*, és a dir, les taules amb la informació amb el programa de BI que s'utilitzarà per a crear el dashboard de la informació, ha estat necessari la instal·lació d'un producte addicional, també gratuït i dels mateixos proveïdors, anomenat *Connector/ODBC*²².

10.2.Qlik Sense



Per a la creació del *dashboard* sobre les dades amb què es compten, s'ha decidit treballar amb un dels softwares líders del mercat de BI: *Qlik Sense*. Tenint en compte les fortaleces i debilitats de *Qlik Sense*, explicades a l'apartat 6.1, en el aquest cas s'ha escollit fonamentalment pels següents motius:

- El preu. Ofereix una versió gratuïta molt completa que disposa de totes les funcionalitats que es necessitaran.
- ETL i emmagatzematge. Incorpora un motor de tractament de dades (*data management*) líder en tot el mercat que en aquest cas, tal i com estan estructurades les dades serà necessari.
- Coneixements previs del software. Es compta amb experiència amb el programa que farà la realització del *dashboard* més accessible i senzill que si es fes amb qualsevol dels altres programes.

²² *Open Database Connectivity*. Concepte que fa referència a poder accedir a qualsevol dada de qualsevol base de dades des de qualsevol aplicació.



11. Implementació a Qlik Sense

Ja s'ha comentat la necessitat imperiosa que tenen les empreses per extreure valor de totes les dades que recopilen, i l'elaboració d'interfícies interactives que mostren de forma entenedora, senzilla i que permeten una ràpida extracció de resultats, és l'opció que la gran majoria de les empreses ha decidit triar, i estan disposades a invertir-hi temps i recursos.

En el cas pràctic d'anàlisi del Churn en què es basa aquest treball, també s'assimila al cas en què es disposa d'una gran base d'informació, difícil d'analitzar en el format tabular original, i es vol transformar en un conjunt de taules i gràfics interactius (dashboards) que mostrin tota la informació útil de la que es disposa i crear-ne valor.

Aquests dashboards pretenen donar suport a tot el conjunt de decisions empresarials, enfocades al tractament de la pèrdua de clients, d'una companyia que es dedica al sector de la venda al detall i que disposa de varies botigues en diferents ciutats, així com un sistema de seguiment de l'activitat dels clients, com targetes clients.

Un cop decidit que el software que s'utilitzarà és Qlik Sense, tot el procés d'elaboració del dashboard es desglossa en tres etapes:

1. Disseny dels dashboards
2. Tractament i transformació de les dades
3. Creació de la interfície interactiva

11.1. Disseny dels dashboards

Abans de posar-se a elaborar un dashboard, es necessita fer la feina prèvia de dissenyar què i com es vol representar, per a qui va destinat (perfil tècnic o de directiu) i si es disposa de les dades necessàries per aconseguir-ho (Taula 12).

Taula 12 Disseny de què es mostrarà, com es mostrarà, i per a qui aniran destinats els dashboards

<p><u>Què</u></p> <p>Què es vol mostrar</p>	<ul style="list-style-type: none"> Visió d'indicadors (KPIs²³) sobre l'evolució econòmica.
--	--

²³ *Key Performance Indicator*. Fa referència a mètriques utilitzades per a quantificar els resultats obtinguts.

	<ul style="list-style-type: none"> • Visió sobre l'evolució dels clients i indicadors relacionats amb ells • Visió sobre l'evolució del Churn • Relació i dependències entre paràmetres, com els ingressos i el nombre de visites
<p><u>Com</u></p> <p>Enfocament dels dashboards. Cada dashboard estarà dissenyat per donar una visió i un detall diferent</p>	<ul style="list-style-type: none"> • Visió global, sense entrar en detalls • Detall a nivell de client • En funció del patró d'activitat del client • Detall a nivell de botiga i posició geogràfica
<p><u>Per a qui</u></p> <p>Per a qui va destinat</p>	<ul style="list-style-type: none"> • Destinat a perfils poc tècnics, amb possibilitat d'entrar en detall si es vol

11.2.Tractament i transformació de les dades (ETL)

Com ja s'ha pogut veure en el punt 9.3, es disposen de varies taules amb tota la informació. Cada taula aporta informació única que no es repeteix en les altres, i per tal de crear i alimentar el dashboard, es necessita unir i enllaçar tota la informació de manera que es tinguin totes les dades en un espai conjunt i unificat. Per exemple, no es disposa de cap taula que mostri a la vegada, per client i període, l'import gastat fins al moment, la predicció del CLV i el nivell d'alarma de que el client abandoni l'empresa. Cada una d'aquestes dimensions es troben en taules diferents: la primera en la ClientChar, el CLV en el ClvOut, i l'alarma en l'AeaOut.

Aquest procés d'extracció, transformació i càrrega de dades a una plataforma, en el nostre cas a Qlik Sense, s'anomena *ETL*. Qlik Sense compta amb un gran motor de tractament i transformació de dades que molts pocs competidors li poden fer ombra, i que en aquest cas, serà de gran utilitat.



11.2.1. Taules finals

Per a l'elaboració definitiva dels dashboards, s'ha basat en tres taules diferents, fruit de modificacions i unions de les originals, cada una amb els camps, existents o no existents en les taules originals, imprescindibles per a la representació de la informació que s'ha dissenyat. Tot i ser taules diferents, estan enllaçades entre elles, de manera que qualsevol valor d'una taula, està directament relacionada amb un altre camp d'una altra taula. Aquí és on intervé la verdadera potència de Qlik Sense, i en general dels softwares de Business Intelligence: la capacitat de carregar grans volums de dades de diferents fonts i mantenir-les enllaçades i relacionades.

El procés de transformació i tractament de les taules originals s'ha dut a terme seguint el següent flux de treball (Figura 6):

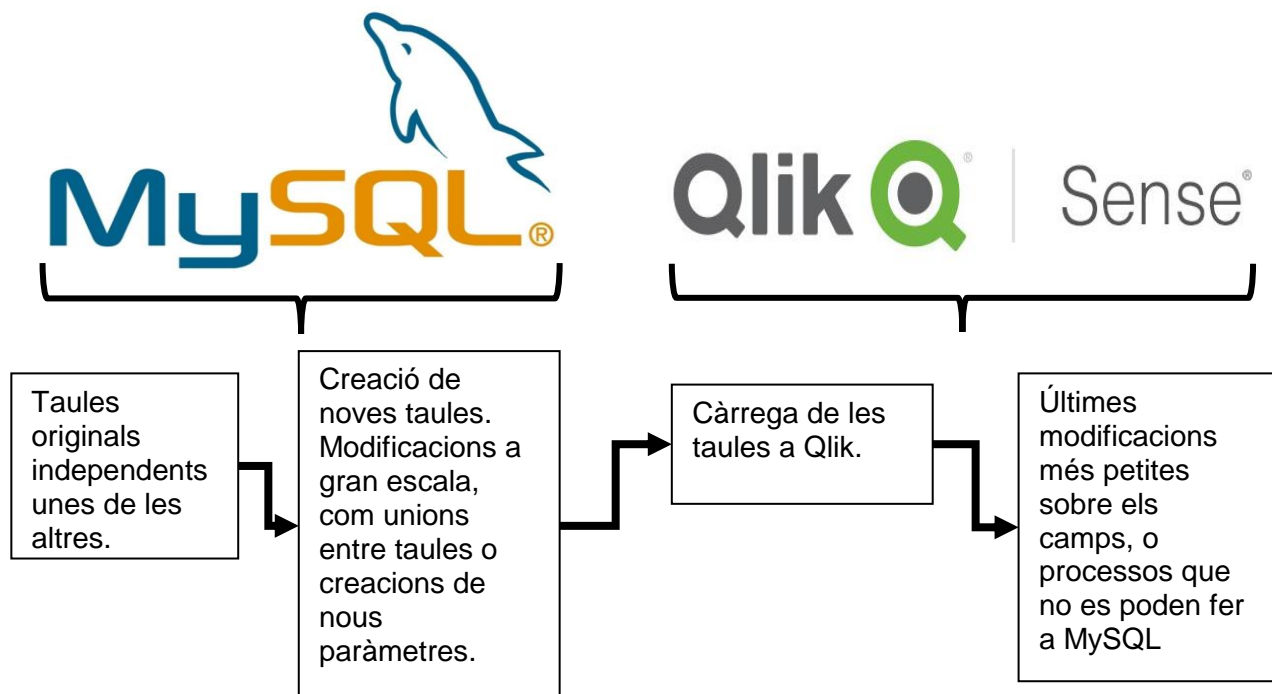


Figura 6 Flux de treball que s'ha dut a terme pel procés de ETL.

El tractament sempre comença a la plataforma *MySQL*, on a partir del llenguatge per tractar bases de dades estructurades: *SQL*, s'han fet totes les transformacions necessàries. Aquestes transformacions han consistit en:

- La creació de noves taules que contenen dimensions que inicialment es trobaven en diferents taules. Aquest procés s'anomena *join*²⁴, i n'hi ha de varis tipus (Figura 7). Donades dos o més taules, es poden combinar per formar-ne una de sola, unint-les per un o varis paràmetres de referència. En el cas en què es troba, es podria combinar les taules *ClientChar* i *AeaOut* donant com a paràmetres de referència el "customerId" i el període ("period").

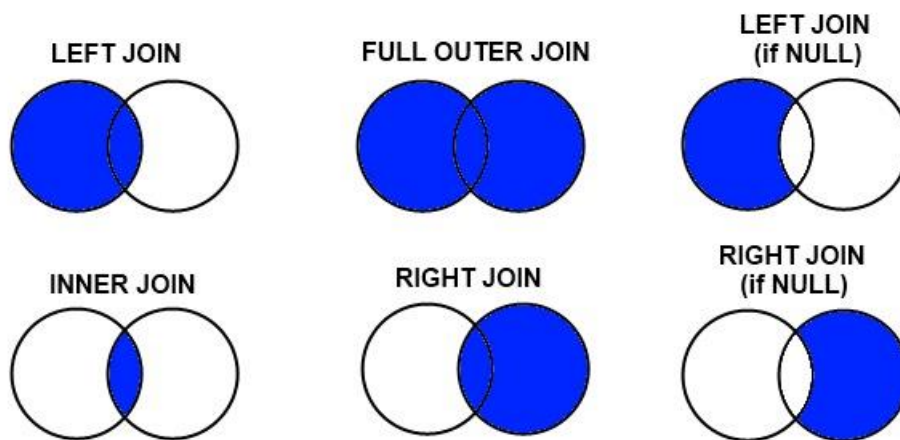


Figura 7 Representació esquemàtica dels diferents tipus de *joins*. Cada cercle representa un taula i la zona colorejada el resultat final de la *join* en qüestió.

- La creació de noves columnes (dimensions) a partir de les que es disposen. La creació de noves dimensions s'han fet per tal de facilitar futurs processos. Un exemple aplicat a aquest cas, seria la creació d'una columna que indiqui si un client, i per cada període, entra en risc de churn sense estar-ho el període anterior, si en canvi en surt, o si es manté en el mateix estat de risc que el període anterior.

11.2.1.1. Taula Base

La taula Base és la taula principal en la que es basa principalment el dashboard. Conté tota la informació recopilada sobre els clients monitoritzats. Aquesta taula és fruit de varies

²⁴ La sentència *JOIN* (en anglès significa unir, combinar) en llenguatge *SQL*, permet combinar registres d'una o més taules d'una base de dades relacional.



JOINS entre les taules: ClientChar (9.3.8), Customer (9.3.7), ClvOut (9.3.10), AeaOut (9.3.9) i Botigues (9.3.1). D'aquesta manera, s'ha unificat en una sola taula tota l'activitat econòmica, informació personal del client, nivells d'alerta de churn, situació geogràfica de les botigues i prediccions, per cada client monitoritzat i període. En definitiva, es té en una sola taula totes les columnes que tenen cada taula per separat. Com algunes dimensions fan referència a la modelització estadística i a paràmetres usats per a analitzar el churn, no s'han carregat a Qlik totes les dimensions directament sinó que se n'ha fet un filtrat, on finalment s'ha quedat amb els següents camps (Taula 13):

Taula 13 Dimensions de la taula final Base

Base			
customerId	mainShop_coord*	regShop_coord*	baseline
period	mainShop_city*	regShop_city*	Revenue
birthDate	secondaryShop	Palive	noOfVisits
Gender	secondaryShop_coord*	CLVpred	activityPattern
Recency	secondaryShop_city*	alarm	
mainShop	regShop	aea	

*Camps nous, creats directament des de l'editor de càrrega de fitxers de Qlik Sense.

Totes les dimensions són dimensions ja existents en les taules d'origen, excepte aquelles amb un "*". En aquest cas, les noves columnes fan referència a informació geogràfica de les botigues. Això ens permetrà més endavant crear un mapa interactiu en un dels dashboards on es representin les posicions de les diferents botigues, i cada una d'elles relacionar-la amb altres dimensions com el nombre de clients o els ingressos aconseguits per cada una.

Totes les taules excepte la taula "Botigues", simplement s'uneixen les columnes pels valors que tenen el mateix *customerId* i *period*, i es parteix de la premissa que la combinació *customerId* + *period*, és única, i no hi haurà dues línies amb la mateixa combinació de dimensions. Per mostrar-ho de manera gràfica (Figura 8), suposant que es parteix de només dos taules, on cada una té les columnes *customerId* i *period*, s'acaba obtenint una de sola, unida pels valors de *customerId* i *period*:

Taula A			Taula B		
customerId	Period	MainShop	customerId	Period	Revenue
290001	1618	1	890004	1618	477.02
890004	1618	23	290001	1618	238.51
50006	1625	1	50006	1625	31.87

+

Join Taula A i Taula B			
customerId	Period	MainShop	Revenue
290001	1618	1	238.51
890004	1618	23	477.02
50006	1625	1	31.87

Figura 8 Representació de la combinació de dues taules a partir de dos columnes de referència

En canvi, la taula “Botigues” s’ha utilitzat com a diccionari, és a dir, una taula de consulta on per cada Id de botiga de la taula “Base”, es busca el/els paràmetres que té associat aquest mateix Id a la taula “Botigues”. Continuant amb l'exemple (Figura 9):

Join Taula A i Taula B			
customerId	Period	MainShop	Revenue
290001	1618	1	238.51
890004	1618	23	477.02
50006	1625	1	31.87

Botigues	
ShopId	City
1	Madrid
23	Barcelona
13	Madrid

Join Taula A i Taula B amb ciutats				
customerId	Period	MainShop	Revenue	City
290001	1618	1	238.51	Madrid
890004	1618	23	477.02	Barcelona
50006	1625	1	31.87	Madrid

Figura 9 Representació de l'ús d'una taula (Botigues) com a diccionari per afegir noves dimensions a una taula original.

En aquest cas, les noves dimensions han estat creades directament des de l'editor de càrrega de fitxers de Qlik. No s'ha fet directament des de MySQL ja que el fitxer que conté la taula “Botigues” no es troba en la mateixa font, sinó que es troba en un fitxer separat en format “.xlsx” (fitxer excel). Tot i tenir dades de diferents fonts d'origen, Qlik Sense permet la càrrega i tractament d'aquesta informació en el seu espai de treball, fet que ens ha



permès relacionar les dades geogràfiques de les botigues.

Finalment, es mostra un esquema (Figura 10) de com es relacionen les diferents taules i quins camps s'utilitzen de cada una d'elles, sense tenir en compte els camps procedents de la taula "Botigues", degut al seu funcionament diferent:

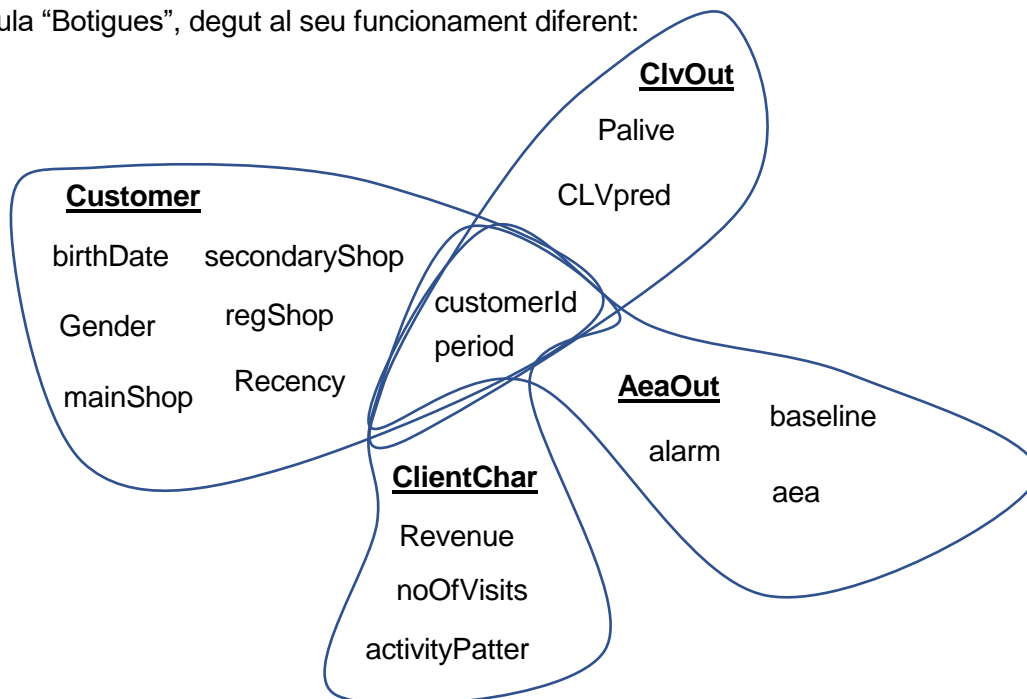


Figura 10 Esquema de les taules utilitzades per a la creació de la taula Base amb les corresponents dimensions. Les que es troben en el mig són les dimensions comunes a cada taula i que s'han utilitzat com a paràmetres d'enllaç.

11.2.1.2. Taula Alarm_CP_NP

La taula Alarm_CP_NP²⁵ té la funció de servir per analitzar l'evolució del churn dels clients. Com ja s'ha comentat anteriorment, per categoritzar si un client es troba en risc d'abandonar l'empresa, s'ha d'analitzar la dimensió "alarm" de la taula "AeaOut" (9.3.9): si el valor de l'alarma és negatiu, es considerarà que es troba en risc de churn per aquell període.

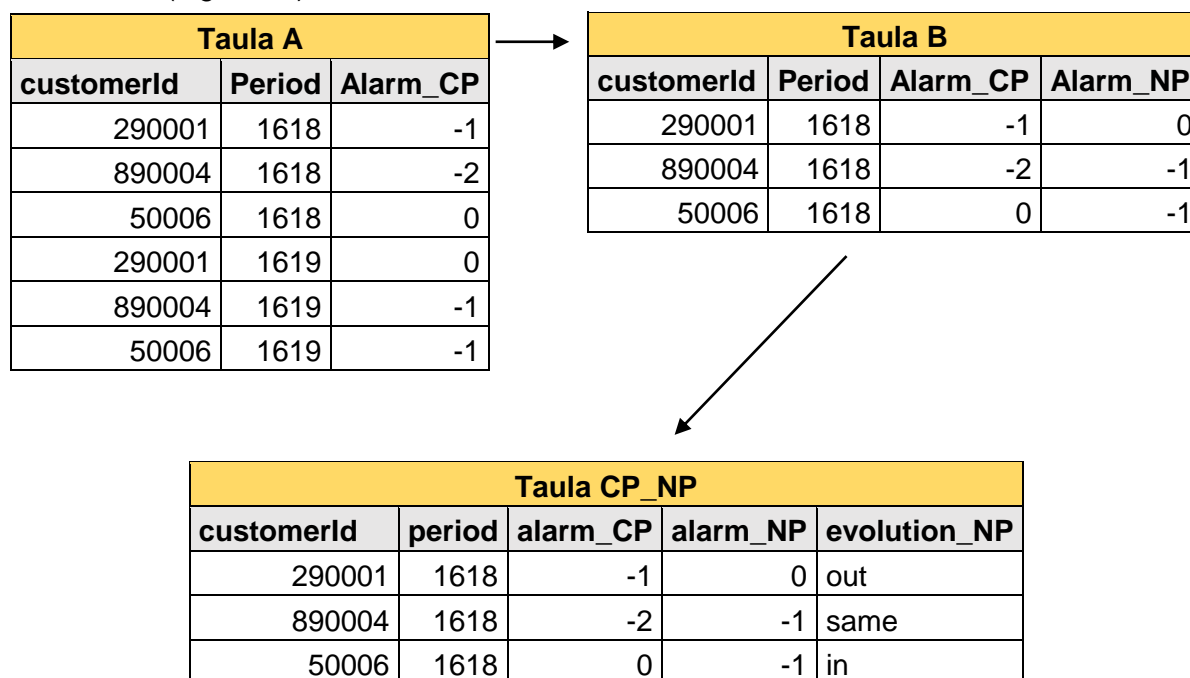
Amb la intenció d'analitzar, en un futur dashboard, quants nous clients entren i surten en risc de churn per cada període, així com el *Customer Lifetime Value* i totes les dimensions associades a aquests clients, s'ha creat una nova dimensió, anomenada "evolution_NP", que indica, per cada client i període, l'evolució del seu nivell d'alarma de cometre churn (Taula 14):

²⁵ CP (*Current Period*): període actual; NP (*Next Period*): següent període

Taula 14 Significat dels diferents valors de la dimensió evolution_NP

evolution_NP	Explicació
in	En el següent període el client passa de no estar en risc de churn a estar-ho.
out	En el següent període el client passa d'estar en risc de churn a no estar-ho.
same	En el següent període el client es manté en l'estat de churn en què es troba en l'actual.

La creació d'aquesta nova taula s'ha basat també en la unió de varies taules, però en aquest cas, s'ha combinat la taula "AeaOut" amb ella mateixa. Primer de tot s'ha hagut de crear una taula que per cada període i client mostri la dimensió "alarm" del període actual, i altre cop la mateixa dimensió però del període següent. Un cop obtinguda aquesta primera taula, la dimensió Alarm_CP_NP consisteix en comparar directament l'alarma del període actual amb la del següent. A continuació es mostra de forma esquemàtica els passos necessaris (Figura 11):

**Figura 11** Esquema del flux de treball utilitzat per a l'obtenció de la dimensió evolution_NP

Així, la taula Alarm_CP_NP està formada per les següents dimensions (Taula 15):



Taula 15 Dimensions que fomen la taula final Alarm_CP_NP

Alarm_CP_NP
customerId
period
alarm_CP
alarm_NP
evolution_NP

11.2.1.3. Taula History

La taula History és una càrrega idèntica de la taula que es té en la base de dades, amb la diferència que s'ha filtrat per només aquells clients que es tenen monitoritzats, ja que són els únics de què es tenen totes les dades, i per tant, apareixen també en les dues taules anteriors carregades a Qlik.

Es necessita carregar la taula History degut a que les seves dimensions són necessàries i úniques, ja que per cada client i període, es mostra el valor puntual (no acumulat, com és el cas de la taula ClientChar) de l'import gastat i el nombre de visites. S'obté, així, la següent taula (Taula 16), la mateixa que en l'apartat 9.3.6:

Taula 16 Dimensions que formen la taula final History

History
customerId
period
shopId
indicatorId
indicatorValue

11.2.2. Relació entre les taules finals

Un cop carregades a Qlik Sense les tres taules finals ja comentades, és el mateix Qlik Sense que s'encarrega de relacionar i enllaçar les diferents taules pels camps necessaris. En aquest cas, els camps que s'utilitzaran d'enllaç, que depenent del cas pot ser un número diferent, són el *customerId* i *period*, ja que són els camps que apareixen a les tres taules. Com ja s'ha comentat anteriorment, la combinació d'aquests dos camps: *customerID+period* és el valor clau, és a dir, no hi ha dos files que pel mateix *customerID+period* presenti les mateixes dimensions.

El mateix motor de càrrega s'encarrega de buscar i crear aquests valors clau. És per això

que un cop es carreguen les taules, Qlik Sense identifica el valor clau de les taules. Com en aquest cas és una combinació de dues dimensions, es crea el que s'anomena una “clau sintètica”, formada per la ja comentada concatenació entre els dos camps. Finalment, Qlik Sense crea el següent esquema del model de dades (Figura 12):

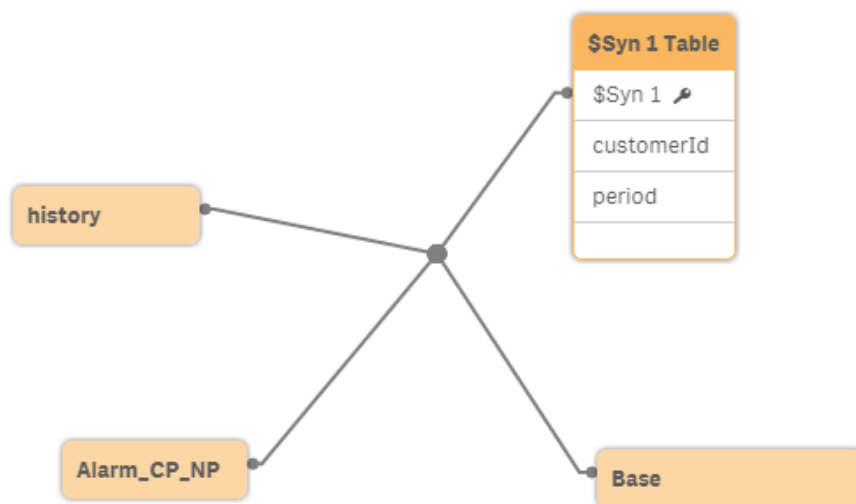


Figura 12 Esquema del model de dades carregat a Qlik Sense. Es veu com les tres taules finals es relacionen entre elles per la clau sintètica *customerID+period*.

11.2.3. Taules addicionals

Amb les dades que s'han carregat a Qlik Sense, es disposa de tota la informació però de només dels clients monitoritzats, que al cap i a la fi, són dels que es té tota la informació que es recull. Per tal d'obtenir indicadors sobre tot el conjunt de clients, sigui quin sigui el tipus de seguiment que se'n tingui (*trackingType*), serà necessari carregar noves taules sense cap filtre per tipologia de client.

Els indicadors que es volen representar són genèrics, com el nombre de clients que es té de cada tipus de client i els ingressos associats a cada un d'ells, sense entrar en més detall.

Per tal fi, tan sols serà necessari carregar de nou, i de forma separada a les taules ja disponibles, la taula sencera de History, i només les dimensions de *customerId* i *trackingType* de la taula Customer. D'aquesta manera, enllaçant aquestes dues taules pel camp de *customerId*, es podrà associar a cada client el tipus de seguiment que se'n té (explicats en la descripció de la taula Customer 9.3.7) i els ingressos i el nombre de visites de cada un i per període.



Degut a que no es volen relacionar aquestes dues taules amb les que ja es tenen carregades a Qlik Sense, s'hauran de carregar amb noms diferents, però el més important és que les dimensions tinguin noms diferents a les dimensions que ja es tenen, ja que sinó, Qlik Sense les enllaçaria creant una clau sintètica amb totes les dimensions que estiguessin repetides. Seguint aquestes condicions, es carreguen les dues taules i les seves dimensions amb els següents noms (Taula 17):

Taula 17 Dimensions de les taules addicionals Wrap_Up_History i Wrap_Up_Customer

Wrap_Up_History ²⁶	Wrap_Up_Customer
customerId_WUP	customerId_WUP
period_WUP	
shopId_WUP	
indicatorId_WUP	
indicatorValue_WUP	trackingType

Les dues taules s'enllacen entre elles pel camp *customerId_WUP*, d'aquesta manera, tots els clients que tinguin algun *indicatorValue_WUP* associat es podrà comptabilitzar. El model carregat a Qlik Sense d'aquestes taules és el següent (Figura 13):

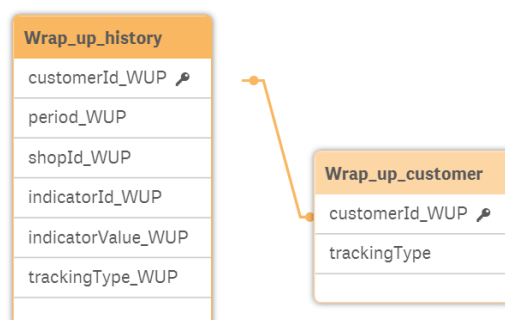


Figura 13 Esquema de les taules addicionals carregades a Qlik Sense, on es veu que s'enllacen entre elles per la dimensió *customerId_WUP*

Finalment, i per tal de poder filtrar i tractar els camps temporals més fàcilment, s'ha aprofitat la potència d'associació de Qlik Sense per crear una nova taula enllaçada a la columna *period*, i que faciliti el tractament d'aquest camp. Com ja s'ha vist, la dimensió *period* consta de quatre dígit, on els dos primers indiquen l'any i els dos últims la setmana de l'any. Per aportar una visió més entenedora i senzilla, i a més a més, afegir noves dimensions d'utilitat com el mes o l'any, s'ha creat una nova columna des de l'editor de càrrega de Qlik Sense a la taula Base, anomenada *period_data*, que simplement transforma el format numèric de

²⁶ *Wrap Up*, en anglès, en aquest cas significa concloure, ja que aquestes taules es faran servir exclusivament per extreure valors de caràcter genèric i global.

period, en format *datetime*²⁷, i s'hi ha enllaçat una nova taula (Taula 18):

Taula 18 Dimensions que formen la taula addicional Date

Date	
period_data	Week
Day	Quarter
Year	MonthYear
Month	WeekYear

A continuació es mostra en un exemple (Taula 19) el significat i la relació entre les diferents dimensions, sent la dimensió origen el *period*, i enllaçant-se amb la taula Base per la columna *period_data*:

Taula 19 Mostra de les diferents dimensions de la taula Date per un *period* donat.

Date								
period	period_data	Day	Year	Month	Week	Quarter	MonthYear	WeekYear
1618	2/5/2016	2/5/2016	2016	maig	18	2	maig 2016	18-2016
1620	16/5/2016	16/5/2016	2016	maig	20	2	maig 2016	20-2016
1627	4/7/2016	4/7/2016	2016	juliol	27	3	juliol 2017	27-2016

Un cop carregades totes les taules, tant les finals com les addicionals, el model de dades sobre aquestes taules carregades a Qlik Sense queda de la següent manera (Figura 14):

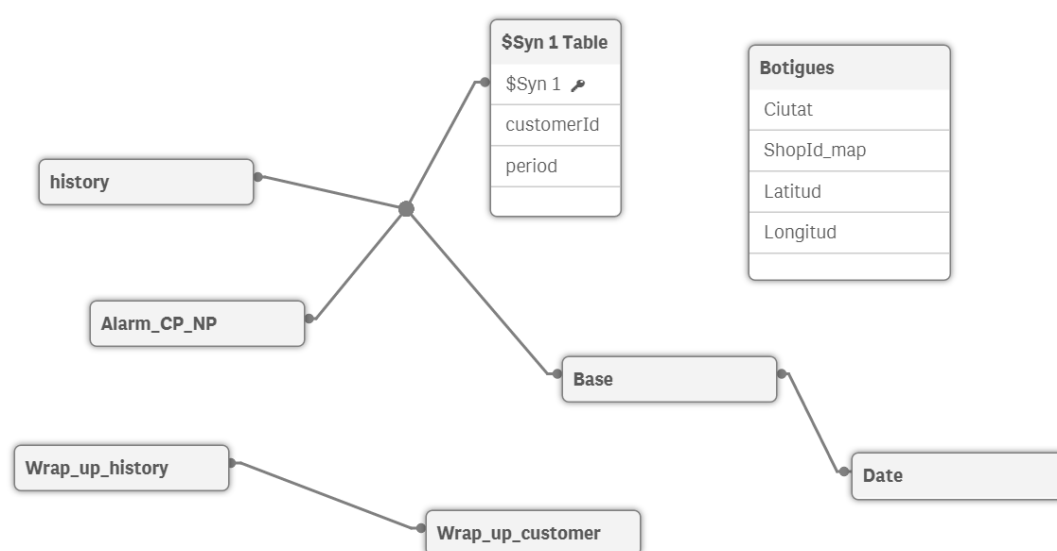


Figura 14 Esquema final del conjunt de taules carregades a Qlik Sense i com s'enllacen entre elles

²⁷ Tipus de format que permet aplicar-hi funcions temporals, com afegir o restar mesos, extreure'n la setmana o l'any, i una multitud d'opcions més.



Es veu com les tres taules principals estan enllaçades entre elles per la clau sintètica *customerId+period*. La taula Base està enllaçada amb la taula Date per tractar les dimensions temporals més fàcilment a través del camp *period_data*, que aquest està directament relacionat amb el camp *period*, i com el camp *period* forma part de la clau sintètica de les tres taules, totes les dimensions de la taula Date també estan relacionades amb les de les taules History, Alarm_CP_NP i Base. Les dues taules addicionals per obtenir dades sobre els clients no monitoritzats, *Wrap_up_history* i *Wrap_up_customer*, estan enllaçades només entre elles a través del camp *customerId_WUP*. Per últim, ens queda la taula Botigues, que és la que conté la informació geogràfica, i aparentment no està enllaçada amb cap altra taula. Tot i així, com ja s'ha comentat anteriorment (11.2.1.1), aquesta taula es fa servir com a diccionari (taula de només consulta) a la taula Base, per associar a tots els Id de botiga, la ciutat i les coordenades d'on està situada la botiga.

11.3. Creació de la interfície interactiva

Un cop realitzada tota la feina prèvia de dissenyar què es vol representar en el dashboard i transformar les dades adequadament per aconseguir tal fi, el pas final és crear la plataforma visual per plasmar tota la informació i els resultats amb el software de Qlik Sense.

Inicialment es crea el que Qlik Sense denomina com a “App”²⁸, que és l'equivalent a dir la carpeta que contindrà tot el conjunt de dashboards. A aquesta App se l'ha anomenada “Anàlisi del Churn”, i a Qlik Sense apareix amb la següent aparença (Figura 15):

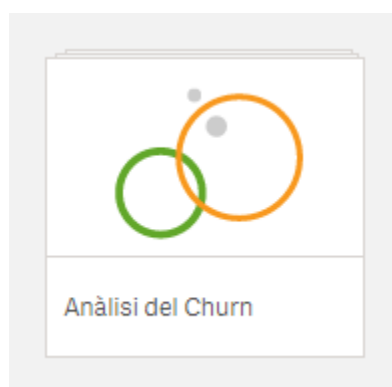


Figura 15 Visió de la App creada a l'interfície de Qlik Sense

Aquesta “App” conté tots els diferents dashboards. S'ha creat un total de 10 dashboards diferents, cada un d'ells responent al disseny que s'havia dissenyat prèviament. L'aparença

²⁸ “App”, de l'anglès *application*, que traduït significa aplicació.

de l'interior de la "App" creada a Qlik Sense amb els diferents dashboards és la següent (Figura 16):

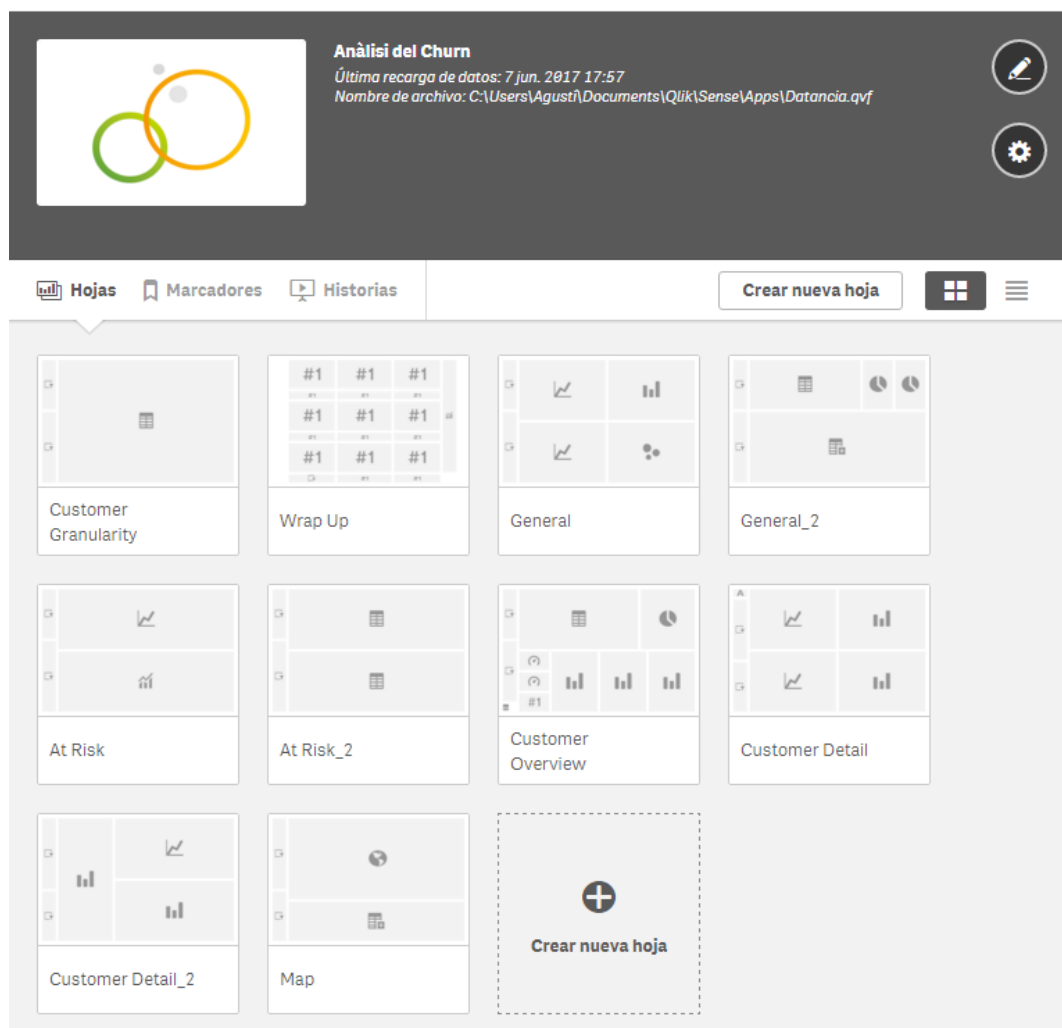


Figura 16 Visió del conjunt de dashboards dins de la App a Qlik Sense



11.3.1. Customer Granularity

Date Filter	cus...	Wee...	perio...	activity...	reve...	noOf...	alar...	CLV...	Pali...	evolut...	Gen...	birth...	mainSh...
WeekYear	252633	25-2016	20/6/2016	Ultra High	5340,68	189	0	4530,41	1	same	M	1947	25
MonthYear	252633	26-2016	27/6/2016	Ultra High	5457,63	193	0	4555,18	1	-	M	1947	25
Quarter	252633	24-2016	13/6/2016	Ultra High	5297,44	186	0	4567,55	1	same	M	1947	25
***	252633	23-2016	6/6/2016	Ultra High	5227,71	184	0	4584,63	1	same	M	1947	25
	252655	19-2016	9/5/2016	High	839,72	149	1	910,58	1	same	M	1946	9
	252655	18-2016	2/5/2016	High	827,2	146	1	915,15	1	same	M	1946	9
	252655	23-2016	6/6/2016	High	973,23	170	1	966,91	1	same	M	1946	9
Customer	252655	24-2016	13/6/2016	High	1001,34	176	1	975,62	1	same	M	1946	9
activityPat...	252655	25-2016	20/6/2016	High	1034,4	183	1	988,34	1	same	M	1946	9
customerId	252655	26-2016	27/6/2016	High	1059,75	186	1	993,39	1	-	M	1946	9
mainShop	252655	20-2016	16/5/2016	High	884,91	154	2	935,82	1	same	M	1946	9
recency	252655	21-2016	23/5/2016	High	918,68	160	2	950,56	1	same	M	1946	9
birthDate	252655	22-2016	30/5/2016	High	956,32	166	2	968,1	1	same	M	1946	9
Gender	252879	22-2016	30/5/2016	Ultra High	8028,01	235	-2	7159,76	0,99	out	M	1946	13
	252879	23-2016	6/6/2016	Ultra High	8243,95	240	0	7227,46	1	same	M	1946	13
	252879	24-2016	13/6/2016	Ultra High	8527,81	244	0	7349,25	1	same	M	1946	13
	252879	25-2016	20/6/2016	Ultra High	9099,19	253	0	7711,69	1	same	M	1946	13

Figura 17 Visió del dashboard Customer Granularity

És el dashboard (Figura 17) que mostra el màxim detall de què es disposa: totes les dimensions per cada client i període.

Consta únicament d'una taula i el panell de filtratge sobre el client i sobre el període. El panell de filtratge és un objecte clau per facilitar la interacció i la visualització de les dades que ens interessen en cada moment. Es diferencien els filtres sobre els clients i els filtres temporals, amb les següents dimensions, ja comentades anteriorment, cada un (Taula 20):

Taula 20 Conjunt de dimensions disponibles en els filtres

Date Filter	Customer Filter
WeekYear	activityPattern
MonthYear	customerId
Quarter	mainShop
Year	Recency
Day	birthDate
Week	Gender
Month	

Aquest panell de filtratge es manté per la resta de dashboards, situant-se sempre a la franja esquerra, tal i com es veu a la figura anterior.

11.3.2. Wrap Up

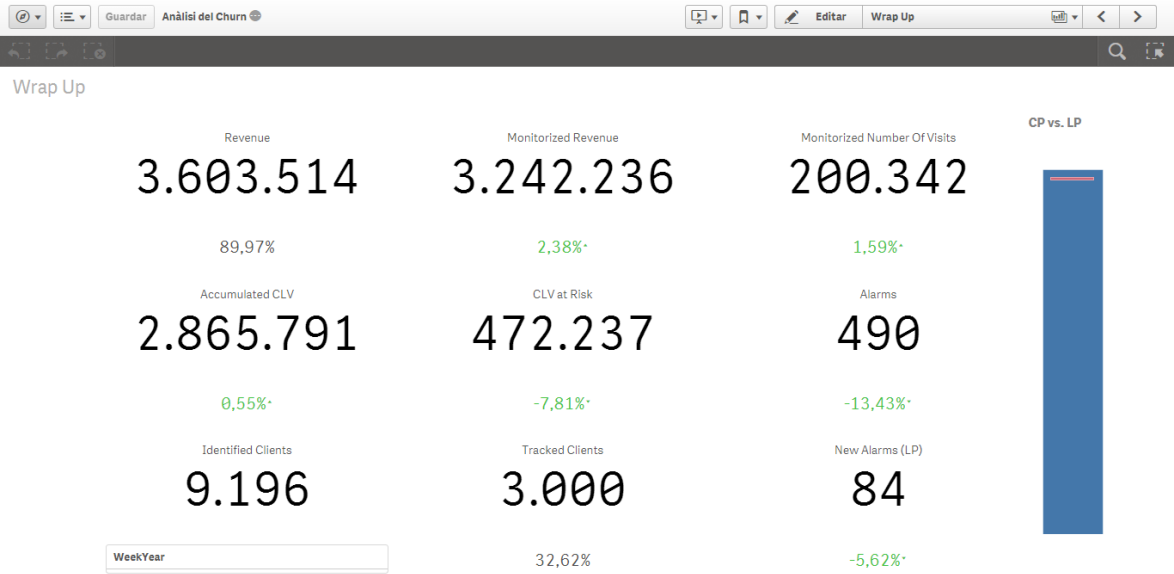


Figura 18 Visió del dashboard Wrap Up

El dashboard (Figura 18) està dissenyat per donar una visió global i ràpida dels indicadors principals sobre l'activitat de l'empresa i el churn (Taula 21). Per defecte els valors fan referència al període més recent:

Taula 21 Significat de cada un dels KPIs que mostra el dashboard

Revenue	Ingressos acumulats fins a l'últim període
Monitorized Revenue	Ingressos acumulats fins al període més recent de només els clients que es tenen monitoritzats
Monitorized Number of Visits	Número de visites acumulades fins al període més recent de només els clients que es tenen monitoritzats
Accumulated CLV	Predicció del CLV del període més recent per als propers 12 mesos
CLV at Risk	De l'anterior valor de la predicció del CLV, només aquell associat a clients amb risc de churn
Alarms	Total d'alarmes negatives en el període més recent, és a dir, total de clients en risc d'abandonar l'empresa
New Alarms	Del total d'alarmes del període, aquelles que són noves
Identified Clients	Total de clients diferents identificats
Tracked Clients	Total de clients monitoritzats

A més a més dels indicadors, també es mostra la variació d'aquests respecte un període de



referència. Per defecte, aquest període és l'anterior al més recent. Els períodes que es volen representar, tant del que es mostraran els indicadors com amb el que es compararà, es poden canviar amb el filtre *WeekYear* de la cantonada inferior esquerra, de manera que l'usuari pot escollir de quin període mostrar els indicadors i respecte quin altre els vol comparar.

A més a més, a la banda dreta del dashboard, es mostra de forma gràfica una comparació ràpida dels ingressos entre els dos períodes que es vulguin: la barra blava representa els ingressos del període base, i la barra vermella els ingressos del període de referència.

11.3.3. General

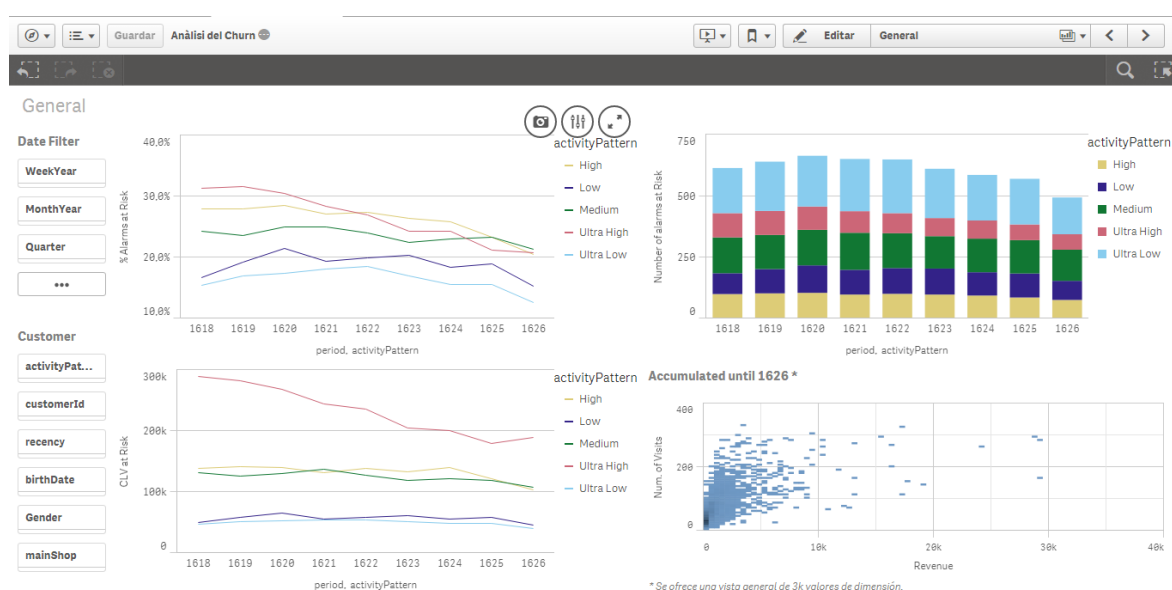


Figura 19 Visió del dashboard General

Aquest dashboard (Figura 19) està dissenyat per mostrar una sèrie de paràmetres de forma totalment gràfica, i per tant, pretén donar a l'usuari una visió ràpida de la situació i permetre identificar tendències i anomalies. Els gràfics mostren, d'esquerra a dreta i de forma descendent:

1. Evolució temporal del percentatge d'alarmes negatives (i per tant, clients en risc de Churn) respecte el total de clients, per cada patró d'activitat diferent (dimensió *activityPattern* de la base de dades).
2. Nombre absolut d'alarmes de cada període i per cada tipus de patró d'activitat.
3. Evolució temporal del CLV en risc i segmentat per cada tipus de patró d'activitat.

4. Relació entre els ingressos i el nombre de visites; ambdós valors acumulats fins al període filtrat (per defecte el període serà el més recent).

A la banda esquerra també es troba el panell de filtratge comú de la resta de dashboards. És important comentar, ja que és el primer dashboard que apareixen gràfics, que Qlik Sense permet maximitzar la vista de cada un dels gràfics i taules en el cas que es desitgi interactuar més detingudament amb una figura en concret.

11.3.4. General_2

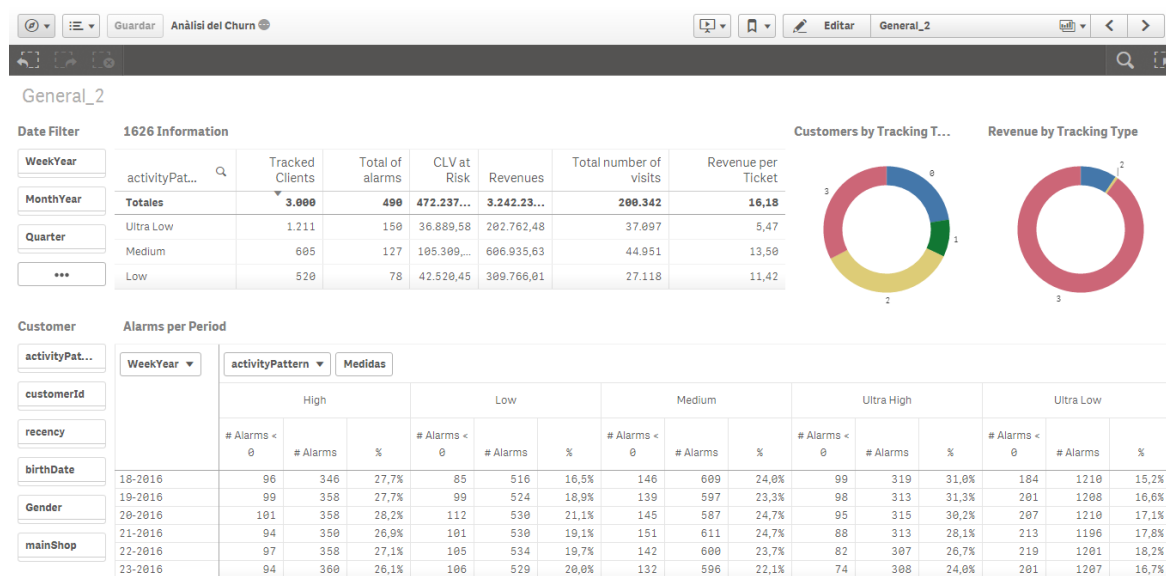


Figura 20 Visió del dashboard General_2

Aquest dashboard (Figura 20) és una continuació de l'anterior. El seu objectiu continua sent de donar a l'usuari una visió general de paràmetres importants i representatius, però en aquest cas, representats de manera més tabular i densa.

A la banda esquerra es té el panell de filtratge, i continuant d'esquerra a dreta i de forma descendent, es té:

1. Taula que per el període indicat (per defecte serà el més recent), indica per cada patró d'activitat del consumidor els següents indicadors: número de clients monitoritzats, número de clients en risc (número d'alarmes), *Customer Lifetime Value* (CLV) en risc, ingressos, numero total de visites i el preu de cada visita.
2. Parella de gràfics circulars. El primer mostra la distribució dels clients segons el tipus de seguiment que se'n tingui, i el segon mostra la distribució dels ingressos també per cada tipus de seguiment.



3. Taula dinàmica on per cada període i tipus de patró d'activitat del client, es mostra el número d'alarmes negatives, el nombre total d'alarmes (per tant, el nombre total de clients monitoritzats) i el percentatge de clients en risc respecte el total.

11.3.5. At Risk

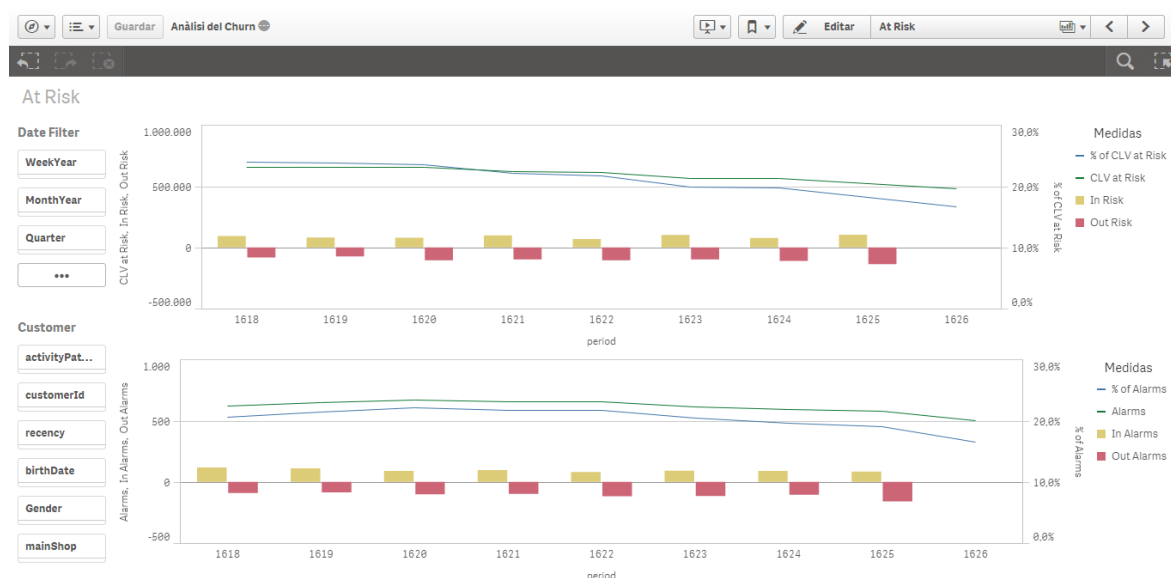


Figura 21 Visió del dashboard At Risk

Aquest dashboard (Figura 21) està dissenyat per representar l'evolució dels clients i el CLV en risc. Al igual que el "General", només conté gràfics, ja que la intenció és que aquest sigui un dashboard intuïtiu i ràpid, per visualitzar tendències i anomalies, i el següent dashboard, "At Risk_2", representi de manera més detallada i densa els valors.

Els dos gràfics que es mostren presenten la mateixa lògica i interpretació, però cada un d'un indicador diferent. L'indicador del de més amunt és el CLV en risc, i el segon indicador és el número d'alarmes (per tant, el número de clients en risc). Per cada període, es representa l'evolució temporal de l'indicador en qüestió (eix de l'esquerra) i el percentatge d'aquest respecte el total (eix de la dreta). A més a més, per cada període s'indica el valor de l'indicador que ha entrat en risc i el que ha sortit durant aquell mateix període. Per exemple, agafant com a indicador el número de clients en risc, si a l'inici d'un període en concret tenim un total de 100 clients en risc, i durant aquest mateix període han entrat 10 clients nous en risc i han deixat d'estar en risc 5 clients, es representaria de la següent manera (Figura 22):

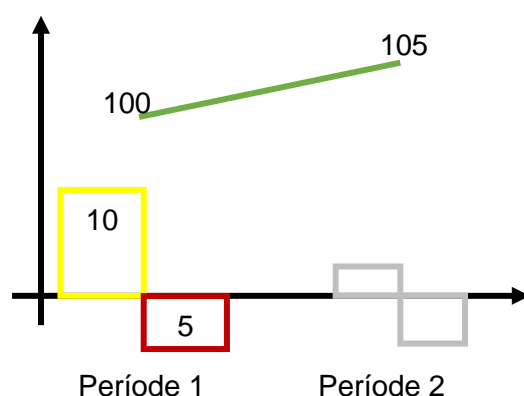


Figura 22 Representació del càlcul dels diferents paràmetres dels gràfics del dashboard

11.3.6. At Risk_2

At Risk_2											
Date Filter 1626 Information											
WeekYear	activityPat...	Tracked Clients	Tracked at Risk	% Tracked at Risk	Revenue	Revenue per Customer	CLV next 12 months	CLV at Risk	CLV at Risk per Customer	% CLV at Risk	Revenue per Ticket
MonthYear	Totales	3.000	490	16,3%	3.242.235,92	1.080,75	2.865.791,49	472.237,49	157,41	16,5%	16,18
Quarter	Ultra Low	1.211	150	12,4%	202.762,48	167,43	272.979,30	36.889,58	30,46	13,5%	5,47
***	Medium	605	127	21,0%	606.935,63	1.003,20	533.215,89	105.309,67	174,07	19,7%	13,50
	Low	520	78	15,0%	309.766,01	595,70	283.565,37	42.520,45	81,77	15,0%	11,42
	High	356	72	20,2%	723.949,45	2.033,57	608.747,65	100.487,25	282,27	16,5%	18,15
Customer	Ultra High	308	63	20,5%	1.398.822,35	4.541,63	1.167.283,28	187.030,54	607,24	16,0%	27,27
activityPat...	Period Granularity										
customerId	WeekYear	CLV at Risk	In Risk	out Risk	% of CLV at Risk	Alarms	In Alarms	Out Alarms	% of Alarms		
recency	Totales	5.234.547,73	696.849,26	811.064,71	20,8%	1.140	638	729	38,0%		
birthDate	25-2016	512.271,41	103.275,96	138.018,42	18,0%	566	84	160	18,9%		
Gender	23-2016	556.692,36	100.260,67	98.072,39	19,8%	607	91	116	20,2%		
mainShop	21-2016	609.000,93	97.966,25	98.050,56	21,9%	647	96	98	21,6%		
	18-2016	642.655,65	92.861,78	82.369,18	23,8%	610	117	91	20,3%		
	19-2016	645.209,78	79.495,11	73.505,86	23,6%	636	110	86	21,2%		
	20-2016	642.816,84	78.861,40	105.148,48	23,3%	600	89	102	22,0%		

Figura 23 Visió del dashboard At Risk_2

El dashboard At Risk_2 (Figura 23) és una continuació de l'anterior, tractant indicadors sobre el risc de churn però de forma més detallada i en forma de taula. És el dashboard que permet a l'usuari observar tots els indicadors en un mateix espai. Com la gran majoria dels dashboards, a l'esquerra es té el panell de filtratge, i la resta de la fulla està ocupada per dos taules a parts iguals. Les taules, per ordre descendent, mostren:

1. Per cada patró d'activitat del client i per un període en concret (per defecte serà el més recent), indica els següents indicadors:



- a. Número de clients monitoritzats
 - b. Número de clients monitoritzats en risc
 - c. Percentatge de clients monitoritzats en risc respecte del total
 - d. Ingressos acumulats
 - e. Ingressos acumulats per client
 - f. Predicció del CLV pels propers 12 mesos
 - g. Predicció del CLV en risc pels propers 12 mesos
 - h. Predicció del CLV en risc per client pels propers 12 mesos
 - i. Percentatge de l'anterior CLV en risc respecte del total
 - j. Import mig gastat pel client per cada visita al supermercat
2. Els valors dels gràfics de la fulla At Risk en forma de taula. Per tant, per cada període es mostra:
- a. Predicció del CLV en risc pels propers 12 mesos
 - b. CLV associat a clients que han entrat en risc durant el període
 - c. CLV associat a clients que deixen d'estar en risc durant el període
 - d. Percentatge del CLV en risc respecte el CLV total
 - e. Número d'alarmes negatives, és a dir, de clients en risc de churn
 - f. Número de clients que han entrat en risc durant el període
 - g. Número de clients que han deixat d'estar en risc durant el període
 - h. Percentatge de clients en risc respecte el total de clients

11.3.7. Customer Overview

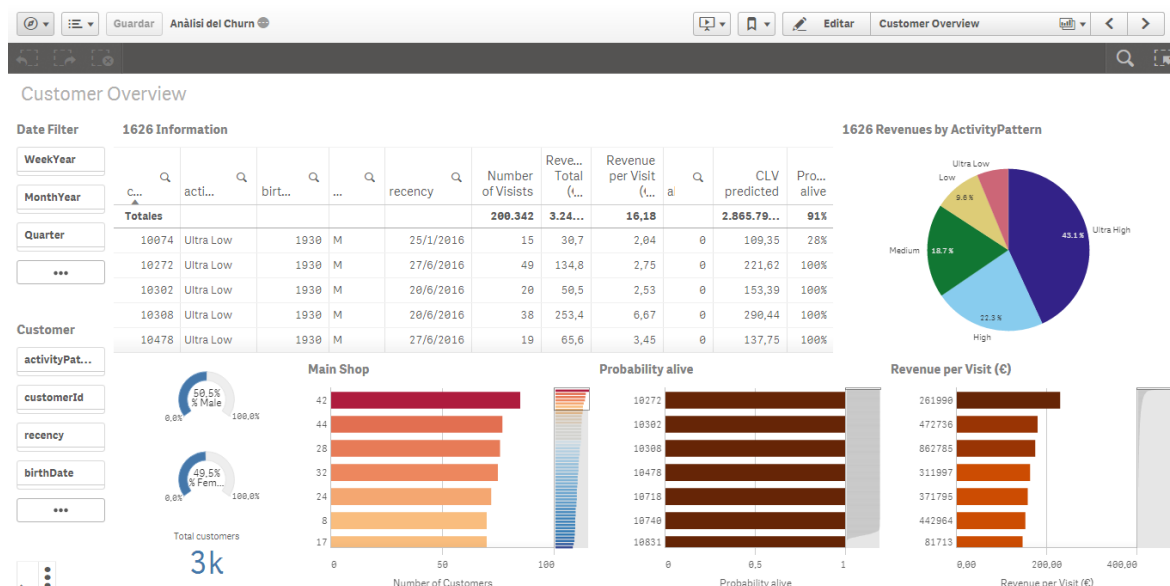


Figura 24 Visió del dashboard Customer Overview

Aquest dashboard (Figura 24) ha estat dissenyat per donar una visió global sobre les característiques dels clients, que si es desitja es pot anar filtrant per les dimensions que es vulgui. Està pensat per analitzar grups de clients, ja que l'anàlisi d'un client en concret en detall es mostra en els dashboards següents: Customer Detail i Customer Detail_2.

El dashboard permet, doncs, obtenir els aspectes més rellevants sobre els clients d'una manera ràpida. La funcionalitat del dashboard és també estudiar comportaments i característiques específiques per grups de clients concrets: segons l'edat del client, la ciutat, la botiga, el gènere, etc.

D'esquerra a dreta i per ordre descendent, el dashboard està compost per:

1. Taula a nivell de client amb les característiques principals de cada un: tipus de patró d'activitat, any de naixement, gènere i data de l'última transacció; i els indicadors de l'activitat principals per cada client durant un període en concret (per defecte el període serà el més recent): número de visites, ingressos, ingressos per visita, nivell d'alarma de churn, predicció del CLV i probabilitat de considerar el client "viu".
2. Gràfic circular que mostra l'aportació als ingressos de cada tipus de patró de l'activitat (*activityPattern*) per un període en concret (per defecte serà el període més recent).



3. Distribució segons el gènere i el nombre total de clients monitoritzats.
4. Distribució de les botigues segons el seu número de clients.
5. Probabilitat, per cada client, de considerar-lo viu.
6. Import gastat per visita de cada client.

11.3.8. Customer Detail

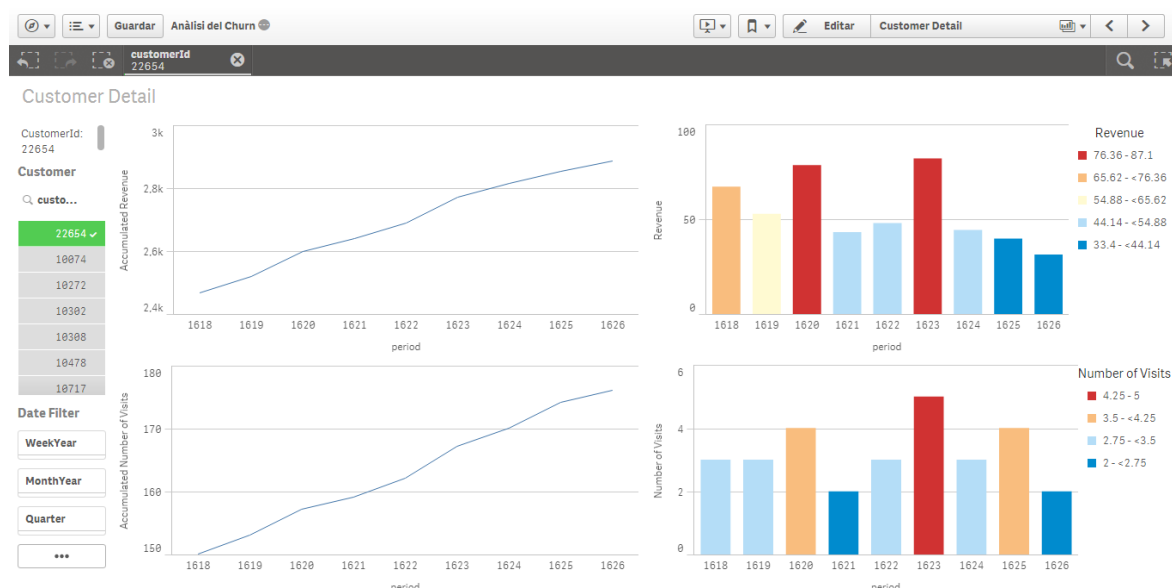


Figura 25 Visió del dashboard Customer Detail

És el primer dels dashboards destinat a representar l'activitat d'un client en concret. Està dissenyat perquè només mostri els gràfics quan es tingui només un client filtrat. La finalitat d'aquest dashboard, al igual que el que ve a continuació (Customer Detail_2), és donar tota la visió disponible des del punt de vista del client (Figura 25).

Els gràfics, d'esquerra a dreta i per ordre descendent, mostren el següent:

1. Ingressos acumulats per cada període
2. Ingressos puntuals de cada període
3. Número de visites acumulades per període
4. Número de visites puntual per període

Per últim, a la banda esquerra, es té com sempre el panell de filtratge, però en aquest cas

els filtres sobre el client s'han reduït, deixant només l'identificador individual del client, deixant però els filtres temporals igual que sempre.

11.3.9. Customer Detail_2

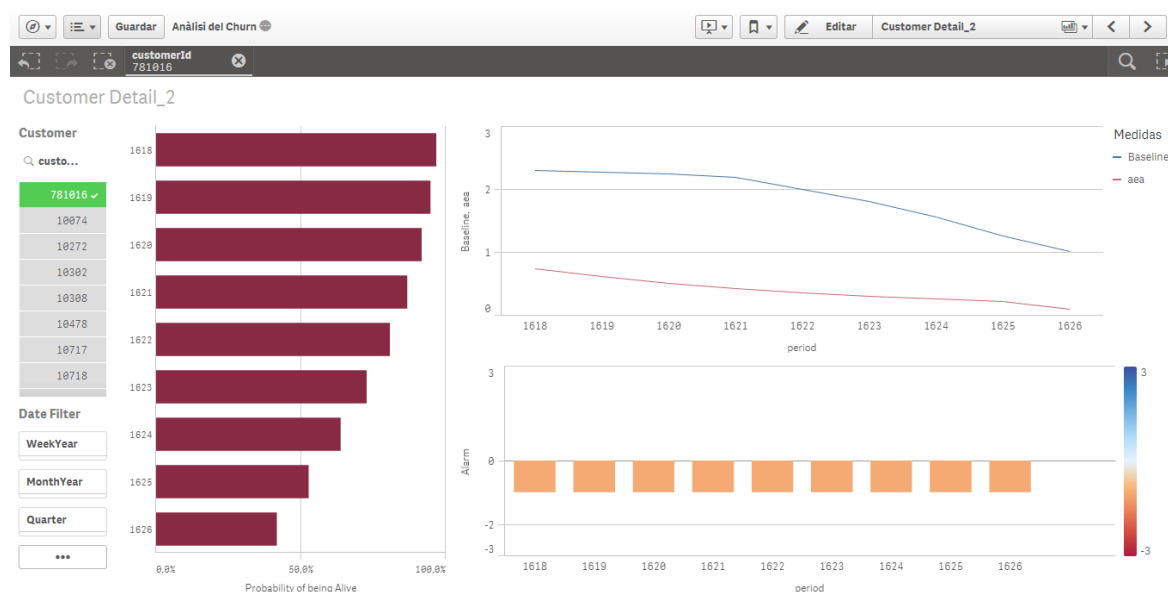


Figura 26 Visió del dashboard Customer Detail_2

És la segona part del dashboard Customer Detail, que es centra en l'activitat del client. Al igual que l'anterior, és un dashboard purament visual i gràfic (Figura 26). Els gràfics que es representen, d'esquerra a dreta i per ordre descendent, són els següents:

1. Gràfic de barres on es mostra l'evolució temporal de la probabilitat de considerar el client viu.
2. Gràfic de línies on es mostra l'evolució temporal dels indicadors *baseline* i *aea*. Es recorda que l'*aea* és un indicador de l'activitat actual del client, i el *baseline* és el valor que s'espera de l'*aea* per mantenir la seva activitat. Per tant, sempre que l'*aea* estigui per sota del *baseline*, el client estarà actuant per sota del que s'espera i el valor de l'alarma disminuirà (indicant potencial de cometre churn), i contràriament, si està per sobre, estarà actuant per sobre del que és esperat i el nivell d'alarma augmentarà.
3. Gràfic de barres sobre l'evolució temporal del valor de les alarmes del client.



11.3.10. Map

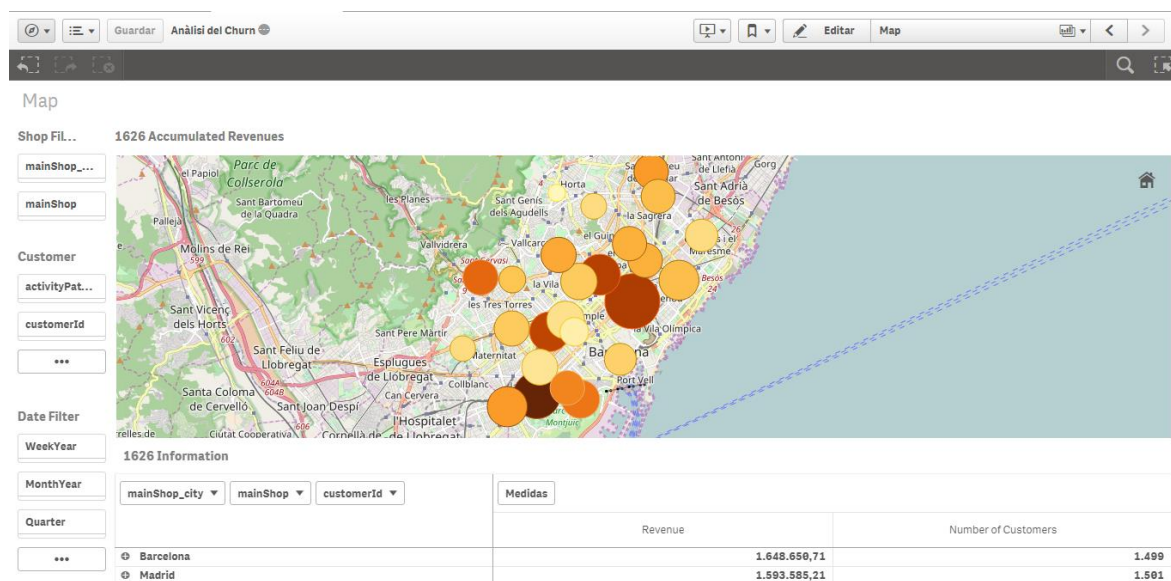


Figura 27 Visió del dashboard Map

És el darrer dels dashboards, i està enfocat a indicadors associats a les botigues, així com la seva representació geogràfica (Figura 27).

El dashboard consta de:

1. El panell de filtratge. A més a més de comptar amb els filtres comuns sobre el client i la data, s'han afegit els filtres relacionats amb la botiga, que són l'Id de la botiga i la seva ciutat.
2. Mapa interactiu on cada punt representa una botiga. A més a més, segons la grandària del punt indica els ingressos de l'establiment, i el color determina el nombre de clients que té.
3. Taula dinàmica on per cada ciutat, Id de botiga o Id de client, mostra els ingressos i el nombre de visites acumulades. Per defecte es representen els valors del període més recent.

11.4. Ús de l'aplicació

Finalment, un cop realitzat tot el desenvolupament dels dashboards, aquesta nova eina de gestió i seguiment ja està llesta per a ser utilitzada i donar suport a la presa de decisions. El conjunt de dashboards dissenyats presenten la informació de tal manera que permet a

l'usuari identificar patrons, anomalies, valors indicatius de l'activitat econòmica, etc., de manera fàcil, ràpida, i sobretot, adaptable a allò que a cada usuari li interessi a cada moment.

Per tal de poder accedir a Qlik Sense i disposar de les seves funcionalitats, entre elles crear i visualitzar els dashboards, es necessita una llicència d'usuari. Existeixen dos tipus d'usuaris: de desenvolupador o simple. El primer tipus d'usuari és el que permet crear i modificar dashboards, mentre que el segon tipus només permet un ús purament funcional i de visualització. D'aquesta manera, des del punt de vista d'aplicació d'un software BI a una empresa, només tindrien llicències de desenvolupadors els tècnics responsables del tractament de les dades i del disseny de dashboards, i tindrien llicències simples tots aquells empleats que necessitessin consultar i analitzar els dashboards.

Un cop creada la "App" amb els dashboards corresponents amb un usuari de desenvolupador, la "App" en qüestió és publicada en l'espai intern de l'empresa a Qlik, de manera que aquells usuaris amb permisos a l'espai on s'hagi publicat podran utilitzar i disposar de totes les funcions dels dashboards.



12. Execució del projecte

El temps de què es disposa per a un projecte és una variable fonamental per tal de decidir-ne la viabilitat. Per això, és indispensable una planificació acurada de les tasques a seguir i, sobretot, contemplar un cert marge per a totes les dificultats i entrebancs que poden alentir l'evolució.

12.1. Tasques principals

Com s'ha evidenciat al llarg de la memòria, aquest treball és de caire teòric i informàtic. És per aquest motiu que gran part de les etapes més bàsiques inclouen aspectes com la programació o la recerca d'informació, a part d'altres com la redacció del document o el disseny de la visualització dels dashboards.

Així, es poden resumir les tasques principals en aquests grans blocs (Taula 22):

Taula 22 Tasques principals del projecte

Tasques	Recerca d'informació	Aprenentatge autònom	Disseny dels dashboards	Programació informàtica	Redacció	Total
% d'hores	18%	10%	7%	40%	25%	100%
Nombre d'hores	72	40	28	160	100	400

Per tant, la primera tasca que s'ha fet per dur a terme el projecte ha estat molta recerca i buscar informació relacionada, principalment, sobre els softwares de BI i sobre el concepte del churn i relacionats a aquest.

Un cop feta tota la recerca per a la part teòrica del treball, el pas següent ha consistit en l'aprenentatge autònom sobre llenguatges de programació i paquets informàtics que faran més endavant. Part de l'aprenentatge ha estat abans de començar el cas pràctic, i inevitablement, una altra part ha estat durant la realització del mateix.

Posteriorment, s'ha hagut de fer la feina prèvia de dissenyar allò que es voldrà mostrar i com es voldrà mostrar. A més a més, una tasca clau ha estat la de decidir quines dades faran falta més endavant per aconseguir mostrar el que es desitja i com transformar les que ja es tenen per obtenir nous paràmetres.

A continuació, i un cop feta la feina prèvia de preparació i disseny, s'ha passat a la pròpia execució del dashboard, pel que han estat necessàries moltes hores de programació informàtica i de buscar solució a un continu de reptes que han anat sorgint.

Finalment, s'ha redactat la memòria explicativa, on s'han unit tots els conceptes teòrics amb la implementació pràctica.

12.2. Planificació detallada

En el següent diagrama de Gantt (Figura 28) es mostra la distribució de les diferents tasques del projecte, havent començat des del mes de Març.

TASQUES	MARÇ		ABRIL		MAIG		JUNY	
Recerca d'informació								
Aprenentatge autònom								
Disseny dels dashboards								
Programació informàtica								
Redacció								

Figura 28 Diagrama de Gantt que mostra el temps requerit per cada tasca del projecte

Com es pot observar, la tasca que ha estat sempre present al llarg del projecte ha estat la recerca d'informació, seguit de la redacció, que s'ha anat fent a mesura que s'avançava en el treball, i de l'aprenentatge autònom, que ha estat imprescindible per adquirir coneixements necessaris per tirar endavant el projecte.

En quant al disseny dels dashboards i la programació informàtica, són tasques relacionades amb el cas pràctic, i per tant, només apareixen en la segona part del projecte. El disseny ha estat suficient fer-lo a l'inici de la part pràctica, mentre que la programació, que ha estat l'etapa que més hores ha necessitat, s'ha anat realitzant de manera continua i seguida.



13. Cost Associat

En aquest projecte, les despeses econòmiques han estat pràcticament negligibles, sobretot en comparació amb altres projectes que precisen de maquinària, materials o tecnologies per a la seva realització. Tot seguit es realitza un estudi per detallar les despeses vinculades al projecte.

Les despeses es divideixen en directes i indirectes. Les despeses directes es desglossen entre les associades als costos humans i les relacionades als costos dels recursos. En canvi, les despeses indirectes s'associen majoritàriament a costos administratius.

13.1.Despeses directes

Com ja s'ha explicat, una part de les despeses directes estan associades a costos humans. Així, tenint en compte la situació del mercat laboral actual, s'ha aproximat el sou d'un enginyer tècnic per hora, i s'ha aplicat al conjunt d'hores invertides (Taula 23):

Taula 23 Cost humà del projecte en funció del preu per hora i les hores totals invertides

Titulació	€/h	Hores dedicades	Cost total
Grau en Enginyeria en Tecnologies Industrials	15	400	6.000 €

L'altra part involucrada a les despeses directes, són els costos associats a recursos materials que s'han utilitzat (Taula 24):

Taula 24 Costos associats a recursos materials

Recurs	Preu
Ordinador portàtil Lenovo Z500	500 €
Llicència paquet Microsoft Office 2016	99 €
Despeses elèctriques	50 €
Total	649 €

D'aquesta manera, el còmput total dels costos directes queda de la següent manera (Taula 25):

Taula 25 Cost total de les despeses directes

Tipus de cost	Pressupost
Humà	6.000 €
Material	649 €
TOTAL	6.649 €

13.2.Despeses indirectes

En aquest projecte, les despeses indirectes es poden associar únicament a costos administratius procedents de la pròpia universitat (matrícula) i de l'impost sobre el valor afegit (IVA), tal i com es mostra a la següent Taula 26:

Taula 26 Despeses indirectes

Motiu	Pressupost
Import pel concepte de Matrícula pel Treball de Final de Grau (IVA inclòs)	474,36
IVA de les despeses directes (21% de 6.649 €)	1.396,29
TOTAL	1.870,65

13.3.Despeses totals

Finalment, sumant el conjunt de despeses directes i indirectes, s'obté el pressupost total que requereix el present projecte (Taula 27):

Taula 27 Detall de les despeses totals del projecte

Tipus de despeses	Pressupost
Directes	6.649
Indirectes	1.870,65
TOTALS	8.519,65



De totes maneres, aquest valor final és estimatiu, ja que s'ha suposat que a l'inici del projecte no es disposava de cap dels recursos, quan realment, al començar el projecte ja es disposava d'ordinador i la llicència del paquet Microsoft Office 2016. Com a conseqüència, tot i els costos calculats, a la pràctica les despeses econòmiques han estat pràcticament nul·les.

14. Impacte ambiental

Valorar els efectes que l'execució d'un determinat treball pot causar al medi ambient abans de realitzar-lo és indispensable per preservar els recursos naturals, i d'aquesta manera, constitueix un requisit bàsic en l'estudi de la viabilitat de qualsevol projecte, degut al problema ambiental que s'està patint arreu del món avui en dia.

No obstant això, degut a que aquest projecte es basa en la recerca d'informació i la utilització de varis softwares (i per tant, de caire informàtic), durant la seva execució no hi ha hagut cap contacte directe amb el medi ambient, i conseqüentment, es considera que el seu impacte ambiental és nul.

Indirectament, es podria considerar que recursos com l'energètic o els fulls de paper han influenciat a la mala preservació del medi ambient, però tenint en compte que la present memòria no supera les 100 pàgines i que el consum energètic ha estat molt baix, es pot afirmar que l'impacte ambiental és gairebé inexistent.



Conclusions

La situació actual de softwares de BI mostra la maduresa del mercat, on les innovacions són cada vegada més constants i els clients demanen més i més prestacions. És evident que hi ha tres softwares que lideren el mercat: Qlik, Tableau i Power BI. Per a un ús estàndard dels softwares qualsevol dels tres és perfectament vàlid i recomanat, però si l'usuari busca un software especialitzat en un camp en concret, com la disponibilitat de gràfics avançats o la transformació de les dades d'origen, llavors no tots els softwares seran igual d'adequats, ja que cada un d'ells presenta una sèrie de punts forts i febles en diferents aspectes.

En quant a la pèrdua de clients com a problema empresarial, s'ha mostrat com la opció més viable i econòmica per mantenir un nombre adequat de clients és la d'aplicar tècniques per evitar la marxa dels clients actuals, en comptes d'invertir exclusivament en intentar adquirir nous consumidors. Les actuals empreses ja evidencien aquest canvi de paradigma aplicant tot un conjunt de polítiques i tècniques comercials destinades a satisfer al client ja existent, i no centrant-se exclusivament en campanyes per adquirir nous consumidors, que no deixarà mai de ser un procés clau i imprescindible pel creixement.

La necessitat de poder parametritzar els clients ha provocat l'aparició de nous conceptes per tal de controlar la seva activitat, la fidelització a l'empresa o el risc de que l'abandoni. En tot aquest procés de control i parametrització, intervé l'habilitat que tingui cada empresa per crear models estadístics predictius, que al cap i a la fi, seran en els que es basaran els resultats, i a més a més, la capacitat de crear eines gràfiques que facilitin la presa de decisions.

Finalment, la creació d'una interfície gràfica i interactiva a través del software de BI de Qlik, ha permès disposar d'aquesta eina de gestió de l'activitat dels clients i de control d'aquells en risc de cometre churn, permetent l'anticipació a aquesta pèrdua i poder aplicar polítiques comercials per tal d'evitar la marxa del client.

L'ús d'una eina de BI és un factor que ha esdevingut clau per poder extreure valor de l'immens volum de dades de què disposen les empreses actualment. Aquests softwares permeten ser competitius en el mercat actual i identificar ràpidament noves oportunitats de negoci i punts anòmals que no es desenvolupen segons l'esperat.

La creació de dashboards no és una feina que es pugui catalogar de senzilla, i és una tasca que està pendent de millora per part dels mateixos proveïdors de softwares de BI perquè cada cop ho siguin més. Requereix la ferma convicció per part de les empreses que aportarà un gran valor, ja que l'elaboració d'aquests eines comporta una inversió de temps i

recursos, que de segur repercutirà en una millora dels resultats i en la presa de decisions.

En global es pot concloure que els resultats del projecte han estat satisfactoris, ja que s'ha aconseguit satisfer els objectius inicials a un cost econòmic i mediambiental pràcticament negligibles, i en el temps predit.



Agraïments

En primer lloc, m'agradaria agrair a la meva tutora del projecte, Sara Fontdecaba Rigat, tota la seva dedicació i implicació, ja que ha estat una peça indispensable del projecte i que gràcies al seu suport el treball ha semblat que fos cosa de dos i no individual.

Per altra banda, voldria agrair també a tot l'equip de Datancia que ha participat i ajudat a tirar endavant el projecte, que de ben segur que sense la seva ajuda no s'hauria pogut aconseguir el resultat obtingut.

Bibliografia

Referències bibliogràfiques

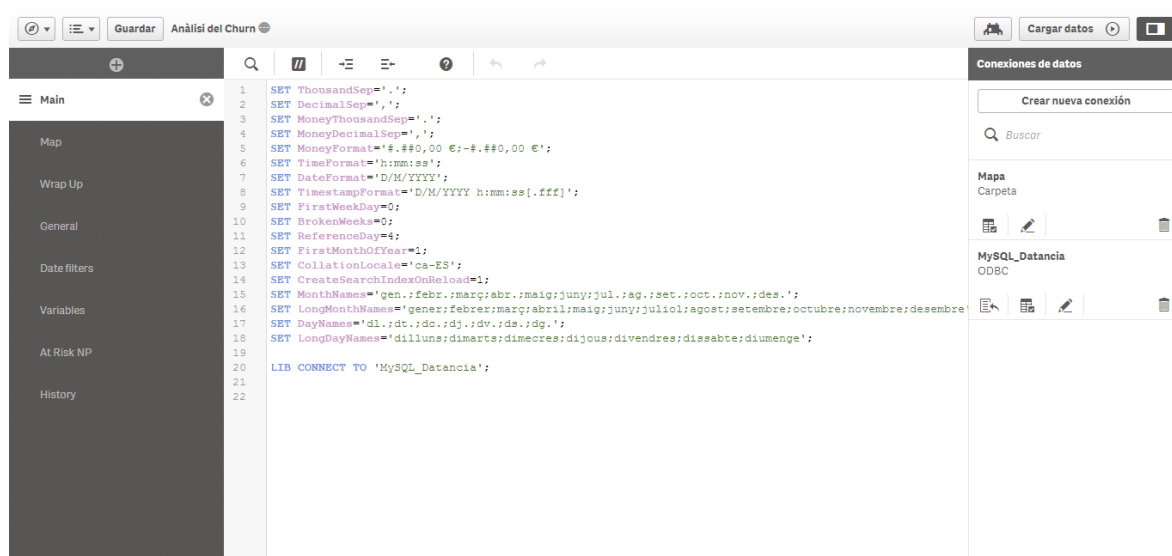
- [1] Xu, L., Tjoa, A. and Chaudhry, S. (2008). *Research and Practical Issues of Enterprise Information Systems II Volume 2*. Guildford: Springer London, Limited
- [2] Gartner.com. (2017). *Technology Research | Gartner Inc.*. [online] Disponible a: <http://www.gartner.com/technology/>.
- [3] Powerbi.microsoft.com. (2017). *Power BI | Interactive Data Visualization BI Tools*. [online] Disponible a: <https://powerbi.microsoft.com>.
- [4] Qlik.com. (2017). *Business Intelligence | Data Visualization Tools | Qlik*. [online] Disponible a: <http://www.qlik.com>.
- [5] Tableau Software. (2017). *Tableau Software*. [online] Disponible a: <https://www.tableau.com/>.
- [6] Rita L. Sallam, Cindi Howson, Carlie J. Idoine, Thomas W. Oestreich, James Laurence Richardson and Tapadinhas, J. (2017). *Magic Quadrant for Business Intelligence and Analytics Platforms*. [online] Gartner.com. Disponible a: <https://www.gartner.com/doc/reprints?id=1-3TYE0CD&ct=170221&st=sb>.
- [7] Datancia.com. (2017). *Datancia – Empower your data*. [online] Disponible a: <http://www.datancia.com>.
- [8] David L. García (2014). *Exploration of Customer Churn Routes Using Machine Learning Probabilistic Models*. UPC Commons.
- [9] Bain.com. (2017). *Global management consulting firm*. [online] Disponible a: <http://www.bain.com>.
- [10] T. Springer, C. Kim, F. Debruyne, D. Azzarello and J. Melton (2014). *Breaking the back of customer churn*. Bain & Company, Inc.
- [11] Mysql.com. (2017). *MySQL*. [online] Disponible a: <https://www.mysql.com/>.



Annexes

A1. Càrrega de les taules a Qlik Sense

La càrrega a Qlik Sense s'ha fet a través del seu editor, on s'hi estipulen totes les càrregues de les diferents taules, principals i addicionals, així com la definició de variables que afavoreixen i faciliten el tractament de la informació. A continuació es mostra la vista de l'editor de càrrega de dades de Qlik Sense, dividida per seccions, i el contingut de cada una de les seccions:



A1.1. Secció Main

Es defineixen paràmetres globals del sistema i es connecta Qlik Sense amb la base de dades que pertoca:

```
SET ThousandSep='.';
SET DecimalSep='.';
SET MoneyThousandSep='.';
SET MoneyDecimalSep='.';
SET MoneyFormat='#.##0,00 €;-#.##0,00 €';
SET TimeFormat='h:mm:ss';
SET DateFormat='D/M/YYYY';
```

```

SET TimestampFormat='D/M/YYYY h:mm:ss[.fff]';
SET FirstWeekDay=0;
SET BrokenWeeks=0;
SET ReferenceDay=4;
SET FirstMonthOfYear=1;
SET CollationLocale='ca-ES';
SET CreateSearchIndexOnReload=1;
SET MonthNames='gen.;febr.;març;abr.;maig;juny;jul.;ag.;set.;oct.;nov.;des.';
SET
LongMonthNames='gener;febrer;març;abril;maig;juny;juliol;agost;setembre;octubre;novem
bre;desembre';
SET DayNames='dl.;dt.;dc.;dj.;dv.;ds.;dg.';
SET LongDayNames='dilluns;dimarts;dimecres;dijous;divendres;dissabte;diumenge';

LIB CONNECT TO 'MySQL_Datancia';

```

A1.2. Secció Map

Es carrega la taula diccionari Botigues.

```

[Botigues]:
LOAD
    Ciutat,
    ShopId as ShopId_map,
    Latitud,
    Longitud
FROM [lib://Mapa/MapaBcnMad.xlsx]
(ooxml, embedded labels, table is Coordenades);

```

A1.3. Secció Wrap Up

Es carreguen les taules addicionals Wrap_Up_History i Wrap_Up_Customer:

```

Wrap_up_customer:
LOAD customerId as customerId_WUP,
    trackingType;

```



```

SQL SELECT customerId,
       trackingType
FROM datancia.customer;

Wrap_up_history:
LOAD customerId as customerId_WUP,
   period as period_WUP,
   shopId as shopId_WUP,
   indicatorId as indicatorId_WUP,
   indicatorValue as indicatorValue_WUP,
   Lookup('trackingType','customerId_WUP',customerId,'Wrap_up_customer') as
trackingType_WUP;
SQL SELECT customerId,
       period,
       shopId,
       indicatorId,
       indicatorValue
FROM datancia.history;

```

A1.4. Secció General

Es carrega la taula Base.

```

Base:
LOAD customerId,
       birthDate,
       sex as Gender,
       recency,
       mainShop,
       GeoMakePoint(Lookup('Latitud','ShopId_map',mainShop,'Botigues'),Lookup('Longitud',
'ShopId_map',mainShop,'Botigues')) as mainShop_coord,
       Lookup('Ciutat','ShopId_map',mainShop,'Botigues') as mainShop_city,
       secondaryShop,

GeoMakePoint(Lookup('Latitud','ShopId_map',secondaryShop,'Botigues'),Lookup('Longitud',
'ShopId_map',secondaryShop,'Botigues')) as secondaryShop_coord,
       Lookup('Ciutat','ShopId_map',secondaryShop,'Botigues') as secondaryShop_city,

```

```

regShop,
GeoMakePoint(Lookup('Latitud','ShopId_map',regShop,'Botigues'),Lookup('Longitud',
'ShopId_map',regShop,'Botigues')) as regShop_coord,
Lookup('Ciutat','ShopId_map',regShop,'Botigues') as regShop_city,
period,
MakeWeekDate('20'&Left(Text(period),2),Right(Text(period),2)) as period_data,
Palive,
CLVpred,
alarm,
revenue,
noOfVisits,
activityPattern,
alarm_1,
alarm_2,
alarm_3,
alarm_4,
alarm_5,
baseline,
aea,
sdAea,
score
;
SQL select c.customerId,cu.birthDate,cu.sex,cu.recency,cu.mainShop,
cu.secondaryShop,cu.regShop,
c.period,c.Palive,c.CLVpred,cl.revenue,cl.noOfVisits,cl.activityPattern,z.*
FROM datancia.clvout c
inner join datancia.clientchar cl
on cl.customerId=c.customerId and cl.period=c.period
left join datancia.customer cu
on cu.customerId=c.customerId
left join (SELECT a.customerId, a.period as last_period,
a.baseline ,a.aea, a.sdAea, a.score, a.alarm,a1.alarm_1,a2.alarm_2,a3.alarm_3,
a4.alarm_4,a5.alarm_5
FROM datancia.aeaout a
inner join (select customerId, period+1 as period, alarm as alarm_1
from datancia.aeaout) a1
on a.customerId=a1.customerId and a.period= a1.period
inner join (select customerId, period+2 as period, alarm as alarm_2
from datancia.aeaout) a2

```

```

on a.customerId=a2.customerId and a.period= a2.period
inner join (select customerId, period+3 as period, alarm as alarm_3
            from datancia.aeaout) a3
on a.customerId=a3.customerId and a.period= a3.period
inner join (select customerId, period+4 as period, alarm as alarm_4
            from datancia.aeaout) a4
on a.customerId=a4.customerId and a.period= a4.period
inner join (select customerId, period+5 as period, alarm as alarm_5
            from datancia.aeaout) a5
on a.customerId=a5.customerId and a.period= a5.period) z
on c.customerId=z.customerId and c.period=z.last_period;

```

A1.5. Secció Date Filters

Es carrega la taula addicional sobre els filtres temporals Date.

```

Date:
Load
    *,
    %KeyDate as Day,
    Year([%KeyDate]) as Year,
    Month([%KeyDate]) as Month,
    Week([%KeyDate]) as Week,
    If(Month([%KeyDate])<=3,1,If(Month([%KeyDate])<=6,2,If(Month([%KeyDate])<=9,3,4)))
as Quarter,
    MonthName([%KeyDate]) as MonthYear,
    Week([%KeyDate])&'-'&Year([%KeyDate]) as WeekYear;
Load
    period_data,
    period_data as %KeyDate
Resident Base;

```

A1.6. Secció Variables

Es defineixen variables que facilitaran l'ús de les dades i la creació futura dels dashboards.

```

let vL.LastPeriod= '=Max(period)';
let lastperiod='=max(period_data)';
let vL.LastLastPeriod= '=Max(period, 2)';
let vL.LastLastLastPeriod = '=Max(period,3)';
set vL.MinPeriod = 1618; //min(period);
let MinPeriod='Min(period)';

```

A1.7. Secció At Risk NP

Es carrega la taula Alarm_CP_NP.

```

Alarm_CP_NP:
Load customerId,
    period,
    CLVpred as CLVpred_CP,
    alarm as alarm_CP,
    CLVpred_NP,
    alarm_NP,
    evolution_NP;
SQL
Select a.customerId,a.period,a.CLVpred,a.alarm,b.CLVpred_NP,b.alarm_NP,
    Case When (b.alarm_NP<a.alarm and b.alarm_NP<0 and a.alarm>=0) Then 'in'
    When (b.alarm_NP>a.alarm and b.alarm_NP>=0 and a.alarm<0) Then 'out'
    else 'same'
    end as evolution_NP
from (select c.customerId,c.period,c.CLVpred,a.alarm
FROM datancia.clvout c,
    datancia.aeaout a
where c.customerId=a.customerId
and c.period=a.period) a
inner join (select c.customerId,c.period-1 as period,c.CLVpred as CLVpred_NP,a.alarm as
alarm_NP
FROM datancia.clvout c,
    datancia.aeaout a
where c.customerId=a.customerId
and c.period=a.period) b
on a.customerId=b.customerId and a.period=b.period
Where a.period >= '$(vL.MinPeriod)';

```

A1.8. Secció History

Es carrega la taula History.

```
history:
LOAD customerId,
    period,
    shopId,
    indicatorId,
    indicatorValue;
SQL SELECT customerId,
    period,
    shopId,
    indicatorId,
    indicatorValue
FROM datancia.history
Where customerId in (Select distinct(customerId) from datancia.aeaout)
and period>= '$(vL.MinPeriod)';
```